

**ANALISIS SENTIMEN PERSEPSI PENGGUNA MYINDIHOME  
MENGUNAKAN METODE *SUPPORT VECTOR MACHINE* (SVM)  
DAN *NAÏVE BAYES CLASSIFIER* (NBC)**

**TUGAS AKHIR**

**Diajukan Sebagai salah Satu Syarat Untuk Memperoleh Gelar Sarjana  
Strata-1 Pada Jurusan Teknik Industri Fakultas Teknologi Industri**



Disusun Oleh:

Nama : Sulton Nur Hakim

No. Mahasiswa : 17 522 019

**PROGRAM STUDI TEKNIK INDUSTRI  
FAKULTAS TEKNOLOGI INDUSTRI  
UNIVERSITAS ISLAM INDONESIA  
YOGYAKARTA**

**2021**

## SURAT KETERANGAN PENELITIAN



FAKULTAS  
TEKNIK INDUSTRI

Gedung KH. Mas Mansur  
Jl. Kaliurang Km 14,5 Yogyakarta  
Telp. (0274) 895287, 898444 ext 2511;  
Fax. (0274) 895007

### SURAT KETERANGAN PENELITIAN

Nomor : 218/A/Ka.Lab DATMIN/FTI-UII/IV/2021

***Assalamu'alaikum Warahmatullahi Wabarakatuh***

Kami yang bertanda tangan dibawah ini, menerangkan bahwa mahasiswa dengan keterangan sebagai berikut :

Nama : Sulton Nur Hakim  
No. Mhs : 17522019  
Dosen Pembimbing : Annisa Uswatun Khasanah, S.T., M.B.A., M.Sc.

Telah selesai melaksanakan penelitian yang berjudul " ANALISIS SENTIMEN PERSEPSI PENGGUNA MYINDIHOME MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM) DAN NAÏVE BAYES CLASSIFIER (NBC)" di Laboratorium Data Mining , Prodi Teknik Industri, Fakultas Teknologi Industri, Universitas Islam Indonesia tercatat mulai tanggal 1 November sampai dengan tanggal 15 Desember 2020.

Demikian surat keterangan kami keluarkan, agar dapat dipergunakan sebagaimana mestinya.

***Wassalamu'alaikum Warahmatullahi Wabarakatuh***

Dikeluarkan : di Yogyakarta

Tanggal : 10 April 2021

Mengetahui,  
Kepala Lab. Data Mining

Annisa Uswatun Khasanah, S.T., M.Sc



## PERNYATAAN KEASLIAN

Demi Allah saya akui bahwa karya ini merupakan karya saya sendiri kecuali kutipan dan ringkasan yang setiap salah satunya telah dicantumkan sumbernya. Jika ditemukan di kemudian hari ternyata terbukti pengakuan saya ini tidak benar dan melanggar peraturan yang sah dalam karya tulis dan hak kekayaan intelektual maka saya bersedia ijazah yang saya terima untuk ditarik oleh Universitas Islam Indonesia.

Yogyakarta, 25 Maret 2021



Sulton Nur Hakim

البعثة الإسلامية الأندلسية

**LEMBAR PENGESAHAN PEMBIMBING**

**ANALISIS SENTIMEN PERSEPSI PENGGUNA MYINDIHOME  
MENGUNAKAN METODE *SUPPORT VECTOR MACHINE* (SVM)  
DAN *NAÏVE BAYES CLASSIFIER* (NBC)**



**TUGAS AKHIR**

**Disusun Oleh:**

**Nama : Sulton Nur Hakim**

**NIM : 17522019**

**Yogyakarta, 25 Maret 2021**

**Dosen Pembimbing**

**Annisa Uswatun Khasanah, S.T., M.B.A., M.Sc.**

**LEMBAR PENGESAHAN PENGUJI**

**ANALISIS SENTIMEN PERSEPSI PENGGUNA MYINDIHOME  
MENGUNAKAN METODE *SUPPORT VECTOR MACHINE* (SVM)  
DAN *NAÏVE BAYES CLASSIFIER* (NBC)**

**TUGAS AKHIR**

Disusun Oleh:

Nama : Sulton Nur Hakim

NIM : 17522019

Telah dipertahankan di depan sidang penguji sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata-1 Teknik Industri

Yogyakarta, 26 April 2021

**Tim Penguji**

Annisa Uswatun Khasanah, S.T., M.B.A., M.Sc.

Ketua

Wahyudi Sutrisno, S.T., M.M.

Anggota 1

Abdullah 'Azzam, S.T., M.T.

Anggota 2

Mengetahui,

Ketua Program Studi Teknik Industri

Fakultas Teknologi Industri

Universitas Islam Indonesia



Dr. Fauzanmawan, S.T., M.M.

## HALAMAN PERSEMBAHAN

*Dengan Menyebut Nama Allah yang Maha Pengasih dan Maha Penyanyang.  
Puji Syukur kehadiran Allah SWT atas nikmat yang telah diberikan. Skripsi ini  
saya persembahkan kepada:*

*Bapak Banu Sodikun dan Ibu Kurotin selaku kedua orang tua saya yang tiada  
henti mendo'akan, mendidik, dan memberi semangat dari kecil hingga saat ini.  
Beliau adalah tokoh utama dan segalanya dalam hidup saya yang tak akan  
tergantikan sampai kapanpun. Terimakasih atas pengorbanan dan segala hal  
yang tidak bisa diungkapkan kata dan tak akan pernah bisa dibalas oleh apapun.*

*Kakak saya Fauzy, adik saya Chavid, dan keluarga besar, terimakasih untuk  
semangat, dukungan, dan motivasi yang selalu diberikan sampai pada tahap ini.*

*Serta kerabat, sahabat, dan teman-teman saya yang selalu membantu dan  
menemani hari-hari dari awal sampai akhir masa perkuliahan ini.*



**MOTTO**

الْوَقْتُ كَالسَّيْفِ إِذَا لَمْ تَقْطَعْهُ قَطَعَكَ

*“Waktu bagaikan pedang. Apabila engkau tidak bisa memanfaatkannya, maka ia akan membinasakanmu”*

(Imam Syafi’i)



## KATA PENGANTAR



*Assalamu 'ailaikum wa rahmatullahi wa barakatuh*

*Alhamdulillahirobbil'alamin*, puji syukur penulis panjatkan kehadiran Allah SWT atas segala rahmat dan hidayah yang diberikan-Nya. Shalawat dan salam senantiasa tercurah kepada Baginda Nabi Muhammad SAW atas petunjuk untuk selalu berada di jalan-Nya sehingga penulis dapat menyelesaikan laporan ini. Laporan Tugas Akhir/Skripsi yang berjudul “ANALISIS SENTIMEN PERSEPSI PENGGUNA MYINDIHOME MENGGUNAKAN METODE *SUPPORT VECTOR MACHINE* (SVM) DAN *NAÏVE BAYES CLASSIFIER* (NBC)” ini merupakan salah satu persyaratan dalam menempuh gelar Sarjana Teknik (S.T.) di Fakultas Teknologi Industri Universitas Islam Indonesia.

Terselesainya Laporan Tugas Akhir/Skripsi ini tidak terlepas dari bimbingan dan dukungan berbagai pihak. Oleh karenanya, pada kesempatan ini dengan segala rasa syukur penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. Ir. Hari Purnomo, M.T. selaku Dekan Fakultas Teknologi Industri Universitas Islam Indonesia.
2. Bapak Muhammad Ridwan Andi Purnomo, S.T., M.Sc., Ph.D. selaku Ketua Jurusan Teknik Industri Fakultas Teknologi Industri Universitas Islam Indonesia.
3. Bapak Taufiq Immawan, S.T., M.M. selaku Ketua Program Studi Teknik Industri Universitas Islam Indonesia.
4. Ibu Annisa Uswatun Khasanah, S.T., M.B.A., M.Sc. selaku pembimbing Tugas Akhir yang telah memberikan bimbingan kepada saya.
5. Kedua orang tua, kakak, dan adik saya untuk semua kasih sayang, perhatian, motivasi, dan doa yang diberikan kepada saya sehingga penyelesaian tugas akhir ini dapat terlaksana dengan lancar.
6. Keluarga Laboratorium *Data Mining* angkatan 2016, 2017, dan 2018, kepala laboratorium Ibu Annisa Uswatun Khasanah, S.T., M.B.A., M.Sc., laboran mas Bayu Hertanta, dan teman-teman yang telah

memberikan doa dan dukungan terutama sahabat seperjuangan Nadhita, Septi, Cut, Amrina, dan Halida.

7. Teman-teman dan sahabat Teknik Industri Fakultas Teknik Industri Universitas Islam Indonesia khususnya angkatan 2017.

Semoga kebaikan yang telah diberikan semua pihak kepada penulis semoga menjadi amal ibadah yang senantiasa dibalas dengan kebaikan yang berlipat oleh Allah SWT. Amiin.

Demikian laporan Tugas Akhir ini, penulis menyadari bahwa masih terdapat banyak kekurangan karena keterbatasan pengetahuan dan kemampuan dalam menyelesaikan laporan ini. Semoga laporan ini dapat memberikan manfaat kepada penulis dan semua pihak yang membaca laporan ini. Amiin.

*Wassalamu 'alaikum wa rahmatullahi wa barakatuh.*

Yogyakarta, 25 Maret 2021



Sul-ton Nur Hakim

## ABSTRAK

Pada era globalisasi seperti saat ini, internet menjadi kebutuhan manusia dalam melakukan berbagai hal. Banyaknya pengguna internet menjadi peluang tersendiri bagi penyedia jasa internet, salah satunya PT Telekomunikasi Indonesia (Telkom). Salah satu produk PT Telkom yaitu *IndiHome*. Sebagai satu-satunya BUMN yang bergerak dibidang telekomunikasi, PT Telkom diharapkan dapat memenuhi kebutuhan masyarakat Indonesia. Namun berdasarkan rating yang didapatkan produk *IndiHome* melalui aplikasi *MyIndiHome* di *Google Play* sebesar 3,5 dari 87.000 lebih ulasan. Hal tersebut yang mendorong dilakukannya penelitian mengenai layanan *IndiHome*. Digunakan data ulasan pada tanggal 1 November 2020 – 15 Desember 2020 dengan jumlah ulasan sebanyak 2.539 ulasan sebagai sampelnya. Data tersebut akan dilakukan pelabelan dan pengklasifikasian dalam kelas sentimen positif dan negatif menggunakan metode *Support Vector Machine* (SVM) dan *Naïve Bayes Classifier* (NBC). Hasil pelabelan setiap sentimen kemudian dianalisis menggunakan metode asosiasi kata untuk menemukan informasi yang dianggap penting dan berguna dalam pengambilan keputusan. Berdasarkan tiga perbandingan data *training* dan *testing* yang dilakukan percobaan sebanyak lima kali disetiap perbandingan, didapatkan rata-rata akurasi dengan metode SVM sebesar 86,54% dan dengan metode NBC sebesar 84,69%. Kemudian, metode asosiasi teks pada kelas sentimen positif di antaranya terkait aplikasi, bagus, membantu, mantap, terimakasih, tagihan, cek, mudah, kecepatan, baik, keren, fitur, internet, dan pelayanan. Sedangkan pada kelas sentimen negatif yang sering dikeluhkan di antaranya terkait aplikasi, bayar, jaringan, wifi, lemot, internet, gangguan, tagihan, pelayanan, mahal, pasang, parah, buruk, lambat, dan lag. Hasil ulasan negatif tersebut dianalisis sebab dan akibatnya menggunakan diagram *fishbone* untuk pemecahan masalah.

Kata kunci: analisis sentimen, *IndiHome*, *Support Vector Machine* (SVM), *Naïve Bayes Classifier* (NBC), asosiasi kata, *fishbone diagram*

## DAFTAR ISI

HALAMAN JUDUL .....	i
SURAT KETERANGAN PENELITIAN .....	ii
PERNYATAAN KEASLIAN .....	iii
LEMBAR PENGESAHAN PEMBIMBING.....	iv
LEMBAR PENGESAHAN PENGUJI.....	v
HALAMAN PERSEMBAHAN .....	v
MOTTO .....	vii
KATA PENGANTAR .....	viii
ABSTRAK.....	x
DAFTAR ISI.....	xi
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR.....	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	5
1.3 Batasan Masalah .....	5
1.4 Tujuan Penelitian.....	5
1.5 Manfaat Penelitian .....	6
1.6 Sistematika Penulisan .....	6
BAB II TINJAUAN PUSTAKA .....	8
2.1 Kajian Induktif.....	8
2.2 Kajian Deduktif .....	14
2.2.1 Jasa Telekomunikasi .....	14
2.2.2 PT Telekomunikasi Indonesia (Telkom) .....	15
2.2.3 <i>Google Play</i> .....	18
2.2.4 <i>Online Consumer Review</i> .....	18
2.2.5 <i>Web Scraping</i> .....	19
2.2.6 <i>Machine Learning</i> .....	20
2.2.7 <i>Text Mining</i> .....	20
2.2.8 <i>Sentiment Analysis</i> .....	22
2.2.9 Klasifikasi .....	23
2.2.10 Pembobotan .....	25
2.2.11 <i>Support Vector Machine</i> .....	25
2.2.12 <i>Naïve Bayes Classifier</i> .....	26
2.2.13 Asosiasi Kata .....	27

2.2.14	Diagram <i>Fishbone</i> .....	27
BAB III METODE PENELITIAN .....		28
3.1	Objek dan Subjek Penelitian.....	28
3.2	Populasi dan Sampel Penelitian.....	28
3.3	Metode Pengumpulan Data .....	28
3.4	Jenis dan Sumber Data .....	28
3.5	Variabel Penelitian .....	28
3.6	Metode Analisis Data .....	29
3.7	Tahap Penelitian .....	30
BAB IV PENGUMPULAN DAN PENGOLAHAN DATA.....		34
4.1	Pengumpulan Data.....	34
4.2	Pengumpulan Data.....	36
4.2.1	Analisis Deskriptif .....	36
4.2.2	<i>Pre-processing Data</i> .....	37
4.2.3	Representasi Model.....	42
4.2.4	Analisis Klasifikasi .....	48
4.2.5	Visualisasi dan Asosiasi.....	57
BAB V PEMBAHASAN.....		68
5.1	Gambaran Umum Persepsi Pengguna <i>IndiHome</i> .....	68
5.2	Hasil Penerapan Metode <i>Support Vector Machine</i> (SVM) dan <i>Naïve Bayes Classifier</i> (NBC) .....	68
5.3	Hasil Klasifikasi dan Asosiasi Kata.....	70
5.4	Usulan Perbaikan Berdasarkan Ulasan Negatif Pengguna .....	72
BAB VI PENUTUP.....		75
6.1	Kesimpulan.....	75
6.2	Saran .....	76
DAFTAR PUSTAKA .....		77
LAMPIRAN.....		83

## DAFTAR TABEL

Tabel 1. 2 <i>Top Brand Index</i> ISP .....	3
Tabel 2. 1 Kelebihan <i>Support Vector Machine</i> dan <i>Naïve Bayes Classifier</i> pada Penelitian Sebelumnya.....	11
Tabel 2. 2 Sejarah PT Telkom .....	16
Tabel 4. 1 Proses <i>Translate</i> .....	38
Tabel 4. 2 <i>Pseudo Code Translate</i> .....	38
Tabel 4. 3 Proses <i>Spelling</i> .....	39
Tabel 4. 4 <i>Pseudo Code Spelling Normalization</i> .....	39
Tabel 4. 5 Proses Mengubah Teks Menjadi Huruf Kecil ( <i>Lowecase</i> ) .....	39
Tabel 4. 6 <i>Pseudo Code</i> Mengubah Besar ke Huruf Kecil.....	40
Tabel 4. 7 Proses Menghapus Angka dan Tanda Baca.....	40
Tabel 4. 8 <i>Pseudo Code</i> Menghapus Angka dan Tanda Baca .....	40
Tabel 4. 9 Proses <i>Tokenizing</i> .....	41
Tabel 4. 10 <i>Pseudo Code Tokenizing</i> .....	41
Tabel 4. 11 Proses <i>Filtering</i> .....	41
Tabel 4. 12 <i>Pseudo Code Filtering</i> .....	41
Tabel 4. 13 Perhitungan Skor Sentimen .....	43
Tabel 4. 14 <i>Pseudo Code</i> Perhitungan Skor Sentimen .....	43
Tabel 4. 15 <i>Pseudo Code</i> Pelabelan Kelas Sentimen .....	45
Tabel 4. 16 Pelabelan Kelas Sentimen Pada Ulasan.....	46
Tabel 4. 17 Perhitungan Skor Sentimen .....	47
Tabel 4. 18 Perbandingan Data <i>Training</i> dan Data <i>Testing</i> 1.....	48
Tabel 4. 19 Perbandingan Data <i>Training</i> dan Data <i>Testing</i> 2.....	49
Tabel 4. 20 Perbandingan Data <i>Training</i> dan Data <i>Testing</i> 3.....	49
Tabel 4. 21 Perbandingan Hasil Akurasi pada Klasifikasi SVM.....	49
Tabel 4. 22 <i>Pseudo Code</i> Proses Klasifikasi Dengan Algoritma SVM.....	50
Tabel 4. 23 <i>Pseudo Code</i> Proses Klasifikasi Dengan Algoritma NBC .....	52
Tabel 4. 24 Perbandingan Akurasi Model Klasifikasi SVM dan NBC .....	55
Tabel 4. 25 <i>Confusion Matrix</i> .....	57
Tabel 4. 26 <i>Pseudocode</i> Proses Visualisasi dan Asosiasi.....	58
Tabel 5. 1 Usulan Pemecahan Masalah .....	73

## DAFTAR GAMBAR

Gambar 1. 1 Penetrasi Pengguna Internet 2019 .....	2
Gambar 2. 1 Logo PT Telkom .....	15
Gambar 2. 2 Portofolio Produk dan Jasa PT Telkom .....	17
Gambar 2. 3 Proses Klasifikasi .....	24
Gambar 2. 4 Proses SVM dalam Menemukan <i>Hyperline</i> .....	26
Gambar 3. 1 <i>Flowchart</i> Penelitian .....	30
Gambar 4. 1 Ekstensi <i>Data Scrapper</i> .....	34
Gambar 4. 2 <i>Website</i> myIndiHome .....	35
Gambar 4. 3 Pembuatan <i>Query</i> .....	35
Gambar 4. 4 Proses <i>Srcaping</i> .....	36
Gambar 4. 5 Hasil <i>Scraping Data</i> .....	36
Gambar 4. 6 Jumlah Ulasan Pengguna <i>myIndiHome</i> di <i>Google Play</i> .....	37
Gambar 4. 7 Diagram Jumlah Ulasan Setiap Kelas Sentimen .....	46
Gambar 4. 8 Pelabelan Kelas Sentimen Setelah Reduksi .....	47
Gambar 4. 9 <i>Barplot</i> Visualisasi Sentimen Positif .....	59
Gambar 4. 10 <i>Wordcloud</i> Ulasan Sentimen Positif .....	60
Gambar 4. 11 Asosiasi Kata Ulasan Sentimen Positif .....	61
Gambar 4. 12 <i>Barplot</i> Visualisasi Sentimen Negatif .....	63
Gambar 4. 13 <i>Wordcloud</i> Ulasan Sentimen Negatif .....	64
Gambar 4. 14 Asosiasi Kata Ulasan Sentimen Negatif .....	65
Gambar 5. 1 <i>Fishbone Diagram</i> .....	72

# BAB I

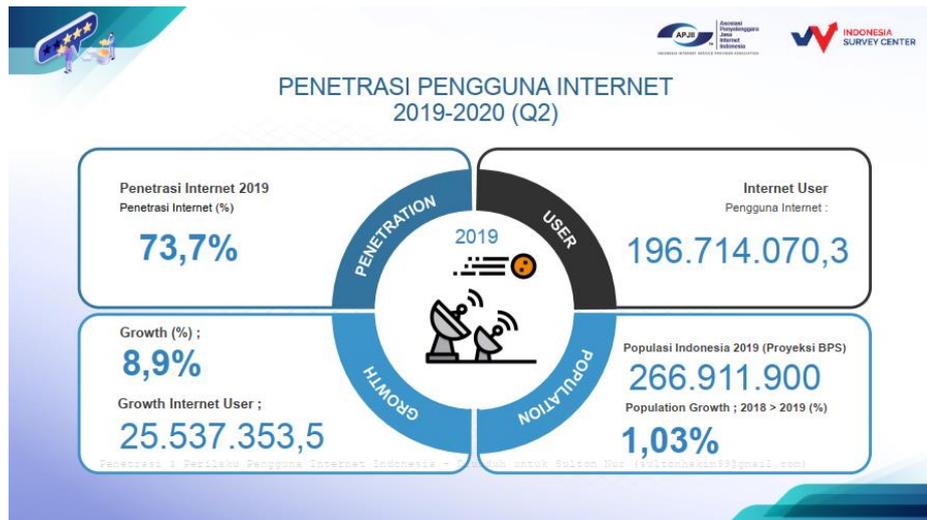
## PENDAHULUAN

### 1.1 Latar Belakang

Era globalisasi tidak dapat dipisahkan dengan perkembangan teknologi. Perkembangan teknologi merupakan hal yang tidak bisa dihindari dalam kehidupan ini. Perkembangan tersebut membentuk pola kehidupan yang serba digital. Adanya berbagai inovasi di bidang teknologi menjadikan masyarakat semakin mudah dalam memenuhi kebutuhannya yang semakin kompleks.

Pada era globalisasi ini, penguasaan teknologi dan internet menjadi prestise dan indikator kemajuan suatu negara (Ngafifi, 2014). Hal tersebut menjadi alasan berbagai negara untuk memberikan perhatian penting dalam perkembangan teknologi dan internet, begitu juga di Indonesia. Perkembangan teknologi dan internet di Indonesia cukup pesat, hal ini ditunjukkan dengan perkembangan teknologi *smartphone*, proses perkembangan jaringan 5G, serta semakin mudah dan luasnya jangkauan internet di Indonesia. Berbagai perkembangan teknologi menjadikan masyarakat semakin mudah dalam mengakses internet. Kemudahan akses internet menjadikan kebiasaan masyarakat berubah yang biasanya mendapatkan informasi melalui media cetak menjadi melalui internet.

Berdasarkan hasil survey Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) pada Juni 2020 menunjukkan bahwa terdapat penetrasi pengguna internet di Indonesia sebesar 73,7% seperti yang tertera pada Gambar 1.1. Nilai tersebut naik sebesar 8,9% dengan total peningkatan sebesar 25 juta lebih pengguna internet. Saat ini, total pengguna internet sebesar 196 juta lebih dari total populasi masyarakat Indonesia sebesar 266 juta jiwa. Jumlah pengguna internet akan terus bertambah seiring dengan perkembangan dan perluasan infrastruktur jaringan di Indonesia (Asosiasi Penyelenggara Jasa Internet Indonesia, 2020).



Gambar 1. 1 Penetrasi Pengguna Internet 2019  
(Sumber: Asosiasi Penyelenggara Jasa Internet Indonesia, 2020)

Banyaknya pengguna internet di Indonesia menjadi peluang pasar yang besar bagi perusahaan-perusahaan penyedia layanan jasa internet. Berbagai perusahaan *provider* jasa internet berlomba-lomba untuk memberikan pelayanan terbaik bagi masyarakat melalui pengembangan layanan internet yang baik dan sesuai dengan kebutuhan masyarakat. Tidak hanya perusahaan swasta, pemerintah juga ikut serta dalam persaingan bisnis dalam sektor telekomunikasi melalui perusahaannya yaitu PT Telekomunikasi Indonesia Tbk (PT Telkom).

Dalam memberikan pelayanan kepada masyarakat, PT Telkom membagi portofolio produk menjadi lima segmen, yaitu *Consumer*, *Enterprise*, *Mobile*, *Wholesale & International Business*, dan segmen “lain-lain” Pembagian ini sebagai upaya dalam menciptakan dan meningkatkan nilai bagi pelanggan melalui bisnis berbasis pada segmen pelanggan atau *Customer Facing Unit* (CFU). Berdasarkan laporan tahunan PT Telkom tahun 2019, salah satu segmen produk yang memiliki peminat cukup banyak adalah pada segmen *consumer*. Segmen ini menyediakan layanan *fixed voice*, *fixed broadband*, IP-TV dan digital dengan total 18,4 juta pelanggan pada tahun 2019. Peminat segmen ini didominasi oleh pelanggan *fixed broadband* pada layanan IndiHome dengan pelanggan sebesar 7 juta customer. Nilai tersebut bertambah menjadi 7,76 juta pelanggan sampai kuartal tiga 2020 (PT Telekomunikasi Indonesia, 2020).

IndiHome merupakan layanan digital menggunakan fiber optik dengan menawarkan layanan Triple Play yang terdiri dari TV interaktif, telepon rumah, dan internet dengan kecepatan akses mencapai 100 Mbps. Dengan kualitas yang

dimiliki, IndiHome menjadi brand *Internet Service Provider* (ISP) paling banyak digunakan selama lima tahun berturut-turut berdasarkan *Top Brand Award*.

Tabel 1. 1 *Top Brand Index* ISP  
Sumber: (Top Brand Award Indonesia, 2020)

BRAND	TOP BRAND INDONESIA (TBI)				
	2016	2017	2018	2019	2020
IndiHome	48,1%	50,3%	42,1%	39,8%	36,7%
First Media	18,6%	17,3%	22,4	29,9%	23,1%
Biznet	4,7%	2,1%	6,4%	8,9%	8,2%
Indosat M2	-	-	-	8,3%	4,5%

PT Telkom menyediakan Plaza Telkom sebagai tempat untuk memenuhi kebutuhan konsumen mulai dari pemasangan baru, keluhan gangguan, hingga pembayaran tagihan layanan IndiHome. Semakin berkembangnya teknologi dan terus bertambahnya pengguna IndiHome, PT Telkom mengembangkan layanan berbasis aplikasi mobile yaitu aplikasi MyIndiHome. Melalui aplikasi ini, *customer* tidak perlu datang ke Plaza Telkom terdekat untuk mendapatkan layanan seperti pemasangan baru, penambahan layanan, informasi tagihan hingga laporan gangguan layanan. Saat ini (20/12/2020) MyIndiHome di *Google Playstore* sudah diunduh sebanyak 5 juta lebih dengan rating 3,6 dan 98 ribu ulasan. Rating yang rendah disertai berbagai ulasan positif dan negatif menunjukkan bahwa pelayanan yang diberikan oleh PT Telkom belum sepenuhnya memenuhi harapan pengguna *IndiHome*. Oleh karena itu perlu adanya evaluasi untuk meningkatkan kualitas pelayanan kepada masyarakat.

Ulasan dari pengguna dapat digunakan sebagai alat yang efektif dan efisien untuk mendapatkan informasi terhadap produk atau layanan yang diberikan. Menurut Fanani (2017) 50% dari pengguna internet bergantung pada rekomendasi *word-of-mouth* (opini) sebelum menggunakan produk, karena *review* dari pengguna lain dapat menyediakan informasi terbaru produk tersebut berdasarkan perspektif pengguna-pengguna lain yang sudah menggunakan produk tersebut. Ulasan di situs *Google Play* akan mempengaruhi citra dari suatu produk/jasa. Citra produk/jasa yang baik akan membentuk opini yang baik pula dari konsumen tentang produk/jasa tersebut, dan diharapkan dapat mendorong terjadinya pembelian produk selanjutnya.

Ulasan mengenai suatu produk/jasa dapat berupa ulasan positif dan negatif. Semakin banyaknya ulasan aplikasi *MyIndiHome* pada situs *Google Play* mengakibatkan perusahaan sulit untuk mendapatkan informasi secara keseluruhan dari ulasan tersebut. Perlu adanya metode yang dapat digunakan untuk mendapatkan informasi pada ulasan secara efektif dan efisien, salah satunya dengan *Text Mining*. Salah satu teknik analisis *Text Mining* yaitu analisis sentimen. Analisis sentimen digunakan untuk mendeteksi keluhan, persepsi produk atau layanan baru, dan persepsi dari suatu merek tertentu. Dengan analisis sentimen perusahaan dapat mengetahui informasi mengenai tanggapan dan sikap dari suatu kelompok atau individu terhadap suatu topik bahasan kontekstual keseluruhan dokumen (Alwasi'a, 2020). Informasi yang didapatkan dari proses tersebut dapat dijadikan sebagai landasan perusahaan untuk melakukan inovasi maupun perbaikan.

Berdasarkan uraian diatas, perlu dilakukan analisis lebih lanjut mengenai ulasan aplikasi *My IndiHome* di situs *Google Play* untuk mengetahui opini pengguna. Selanjutnya opini akan diklasifikasikan pada ulasan positif atau negatif. Evaluasi ulasan tersebut diolah menggunakan analisis sentimen menggunakan pendekatan klasifikasi. Terdapat dua metode yang akan digunakan, yaitu metode *Super Vector Machine* (SVM) dan *Naïve Bayes Classifier* (NBC). Metode klasifikasi *Super Vector Machine* (SVM) dipilih karena mampu menemukan *hyperline* terbaik sebagai pemisah (Ubaidillah, et al., 2019), memiliki akurasi yang baik dalam klasifikasi, proses *learning* yang cepat (Budianto, et al., 2018), dan metode klasifikasi teks yang paling efektif karena dapat mengelola ruang fitur yang besar dan mempunyai kemampuan generalisasi yang tinggi (Dharmadhikari, et al., 2011). Hal tersebut terlihat pada hasil penelitian Steven (2020) yang menunjukkan metode SVM memiliki nilai *Area Under Curve* (AUC) terbesar dibandingkan metode *Neural Network*, *K-Nearest Neighbour*, dan *Decision Tree* dalam mengklasifikasikan ulasan pariwisata Indonesia pada media sosial *Twitter*. Sedangkan metode *Naïve Bayes Classifier* (NBC) merupakan metode paling sederhana dari pengklasifikasian probabilitas dan namun memiliki tingkat akurasi yang sangat tinggi ketika diaplikasikan pada *database* dengan *bigdata* (Wati, 2016). Hal tersebut terlihat pada hasil penelitian Baid, et al (2017) yang menunjukkan bahwa metode NBC memiliki akurasi tertinggi dibandingkan metode 81,45%, dibandingkan akurasi

*K-Nearest Neighbour* dan *Random Forest* dalam mengklasifikasikan *review* film pada situs IMDb.

Setelah proses klasifikasi, dilakukan ekstraksi dan eksplorasi informasi pada setiap sentimen positif dan negatif menggunakan analisis deskriptif dan asosiasi kata untuk mendapatkan kata yang sering muncul pada ulasan pengguna *IndiHome*. Pada sentimen negatif akan dilakukan analisis menggunakan diagram *fishbone* untuk mengetahui faktor-faktor penyebabnya, sehingga hasilnya dapat digunakan PT Telkom menjadi landasan dalam melakukan perbaikan layanan. Penelitian ini diharapkan dapat menyajikan klasifikasi analisis sentimen yang baik dan tepat sehingga dapat memberikan informasi yang berguna untuk evaluasi kinerja pelayanan *IndiHome*.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang ada, permasalahan yang akan dikaji dalam penelitian ini adalah: “Bagaimana analisis sentimen ulasan aplikasi *myIndiHome* pada situs *Google Play* menggunakan metode *Support Vector Machine* (SVM) dan *Naïve Bayes Classifier* (NBC)?”.

## 1.3 Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Data yang digunakan merupakan data ulasan mengenai aplikasi *myIndiHome* pada situs *Google Play* yang diambil pada tanggal 1 November 2020 – 15 Desember 2020.
2. Ulasan yang diambil merupakan ulasan yang menggunakan bahasa Indonesia.

## 1.4 Tujuan Penelitian

Tujuan dalam penelitian ini adalah untuk mengetahui sentimen ulasan aplikasi *myIndiHome* pada situs *Google Play* menggunakan metode *Support Vector Machine* (SVM) dan *Naïve Bayes Classifier* (NBC).

## 1.5 Manfaat Penelitian

Manfaat pada penelitian ini adalah sebagai berikut:

1. Mengetahui gambaran umum mengenai persepsi pengguna *myIndiHome* berdasarkan data ulasan pada website *Google Play*.
2. Hasil klasifikasi sentimen dapat dijadikan sebagai bahan evaluasi kepada pihak manajemen PT. Telkom terhadap kepuasan pengguna layanan *IndiHome* yang berguna untuk pengembangan pelayanan jaringan internet di Indonesia.
3. Memudahkan pihak yang memiliki kepentingan untuk melihat informasi yang terdapat pada ulasan atau komentar yang sangat banyak, sehingga dapat fokus melakukan penanganan serta evaluasi kearah yang lebih baik.

## 1.6 Sistematika Penulisan

Sistematika penulisan dibuat agar penulisan laporan penelitian dapat tersusun rapi dan terstruktur. Berikut merupakan sistematika penulisan pada penelitian ini:

### **BAB I           PENDAHULUAN**

Membuat kajian singkat tentang latar belakang permasalahan, perumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penelitian.

### **BAB II          KAJIAN LITERATUR**

Berisi konsep dan prinsip dasar untuk memecahkan masalah penelitian. Disamping itu juga untuk memuat uraian tentang hasil penelitian yang pernah dilakukan sebelumnya oleh peneliti lain yang memiliki hubungan dengan penelitian yang dilakukan.

### **BAB III        METODE PENELITIAN**

Mengandung uraian mengenai populasi dan sampel penelitian, jenis dan sumber data, variabel penelitian, metode pengumpulan data, metode analisis data, dan diagram alir penelitian.

### **BAB IV        PENGUMPULAN DAN PENGOLAHAN DATA**

Berisi tentang data yang didapatkan selama penelitian dan proses menganalisa data tersebut. Hasil pengolahan data ditampilkan baik dalam tabel maupun grafik. Pada sub bab ini merupakan

acuan dalam pembahasan hasil yang akan ditulis pada sub bab pembahasan hasil.

## **BAB V        PEMBAHASAN**

Melakukan pembahasan hasil yang diperoleh dalam penelitian, dan kesesuaian hasil dengan tujuan penelitian sehingga dapat menghasilkan sebuah rekomendasi.

## **BAB VI        PENUTUP**

Berisi tentang kesimpulan terhadap analisis yang dibuat dan rekomendasi atau saran-saran atas hasil yang dicapai dan permasalahan yang ditemukan selama penelitian.

## **DAFTAR PUSTAKA**

## **LAMPIRAN**



## BAB II

### TINJAUAN PUSTAKA

#### 2.1 Kajian Induktif

Penelitian yang dilakukan oleh Rofiqoh, et al. (2017) mengenai analisis sentimen tingkat kepuasan pengguna penyedia layanan telekomunikasi seluler Indonesia pada *twitter* menggunakan metode *Support Vector Machine* dengan *Lexicon Based Features* sebagai pembaharuan fiturnya. Hasil dari penelitian ini menunjukkan bahwa analisis sentimen dengan metode *Support Vector Machine* dan *Lexicon Based Features* sebesar 79%, sedangkan sistem analisis sentimen tanpa menggunakan *Lexicon Based Features* menghasilkan akurasi 84%.

Fanissa1, et al. (2017) dalam penelitiannya menggunakan metode *Naïve Bayes* dengan seleksi fitur *Quaery Expansion Ranking* untuk mengurangi kompleksitas komputasi pada proses klasifikasi. Data yang digunakan diambil dari ulasan-ulasan pada *website TripAdvisor*. Hasil penelitian ini menunjukkan bahwa hasil seleksi fitur sebesar 75% dan memiliki akurasi terbaik sebesar 86,6%.

Penelitian yang dilakukan oleh Muthia (2017) menggunakan *Naïve Bayes* yang dikombinasikan dengan *Genetic Algorithm* untuk meningkatkan akurasinya. Hasil dari penelitian ini menunjukkan bahwa akurasi yang didapatkan dengan menggunakan *Naïve Bayes* sebesar 86,5%. Nilai akurasi tersebut meningkat menjadi 90,5% ketika metode *Naïve Bayes* dikombinasikan dengan metode *Genetic Algorithm*.

Rahmawati, et al. (2017) dalam penelitiannya menggunakan metode *Support Vector Machine* dan *K-Means Clustering* untuk mengklasifikasikan respon masyarakat mengenai pelaksanaan pilkada serentak. Data yang digunakan berjumlah 3000 *tweet* Bahasa Indonesia. Dari hasil proses klasifikasi didapatkan metode SVM memiliki nilai akurasi lebih tinggi dibandingkan metode *K-Means Clustering* yaitu sebesar 91%.

Baid, et al. (2017) melakukan penelitian untuk mengetahui analisis sentimen mengenai *review* film pada situs IMDb dengan membandingkan beberapa algoritma yaitu *Naïve Bayes*, *K-Nearest Neighbour*, dan *Random Forest*. Pengolahan data pada penelitian ini menggunakan *software Weka*. Hasilnya menunjukkan bahwa algoritma *Naïve Bayes* memiliki akurasi terbaik

dengan nilai sebesar 81,45%, dibandingkan akurasi algoritma *K-Nearest Neighbour* sebesar 55,30%, dan *Random Forest* sebesar 78,65%.

Penelitian yang dilakukan Permana, et al. (2017) memiliki tujuan untuk mengklasifikasikan ulasan mengenai kepuasan siswa terhadap lembaga konsultan pendidikan dengan menggunakan metode *Naïve Bayes*. Data yang digunakan didapatkan dari data ulasan di media sosial *Twitter*. Hasil dari penelitian menunjukkan bahwa algoritma *Naïve Bayes* memiliki akurasi sebesar 84%.

Song, et al. (2017) melakukan penelitian tentang klasifikasi ulasan novel berdasarkan data ulasan pembaca di sosial media *Twitter*. Metode penelitian yang digunakan adalah *Multinomial Naïve Bayes* (MBC), *Maximum Entropy*, dan *Support Vector Machine* (SVM) dengan data sebanyak 1.600.000 *tweets*. Terdapat banyak faktor yang mempengaruhi performansi, seperti koleksi data, banyaknya fitur, metode pra-proses, dan pemilihan algoritma klasifikasi. Hasil penelitian menunjukkan bahwa akurasi dipengaruhi oleh jumlah data dan pemilihan algoritma klasifikasi. Akurasi terbaik ditunjukkan oleh metode MBC dengan akurasi sebesar 85,33% dibandingkan dengan metode yang lain.

Haryanto, et al. (2018) dalam penelitiannya menggunakan metode *Support Vector Machine* berderajat dua dan *Query Expansion*. Data yang digunakan berjumlah 400 data komentar yang didapatkan dari situs Tokopedia. Berdasarkan hasil pengujian, didapatkan akurasi metode *Support Vector Machine* dan *Query Expansion* sebesar 96,25% dan akurasi menggunakan metode *Support Vector Machine* tanpa *Query Expansion* sebesar 94,75% .

Dalam penelitian yang ditulis Al-Smadi, et al. (2018), dilakukan analisis sentimen pada data *review* hotel di Arab. Terdapat perbandingan antara dua metode pengklasifikasian, yaitu *Recurrent Neural Network* (RNN) dan *Support Vector Machine* (SVM) dengan dataset berjumlah 24.028 *review*. Hasil penelitian menunjukkan bahwa metode SVM memiliki performa yang lebih baik dibandingkan RNN berdasarkan nilai F-1 dan akurasi dengan nilai sebesar 89,9% dan 95,4%.

Bordoloi & Biswas (2018) dalam penelitiannya membandingkan metode *Naïve Bayes* (NB), *Support Vector Machine* (SVM) dan *Maximum Entropy* pada analisis sentimen produk. Data ulasan produk didapatkan pada situs *Amazon* dan *Flipkart*. Terdapat tiga dataset produk yang digunakan, yaitu dataset ulasan

produk *Lenovo K3*, *Redmi Note 1*, dan *Iphone 5* dengan jumlah 1500, 1500, dan 1750 data ulasan. Hasil penelitian menunjukkan bahwa metode *Naïve Bayes* memiliki akurasi tertinggi dibandingkan dua metode lainnya pada ketiga *dataset* yang digunakan.

Alizah, et al. (2020) dalam penelitiannya membandingkan metode *Naïve Bayes* dan *Support Vector Machine* untuk mengklasifikasikan komentar masyarakat pada sosial media *Twitter* mengenai “*lockdown*”. Data yang digunakan merupakan *live tweet* pada 19 April 2020 pada pukul 12.42 sampai 15.02. Hasil penelitian ini menunjukkan bahwa metode *Support Vector Machine* memiliki akurasi sebesar 87%, sedangkan *Naïve Bayes* sebesar 81%.

Ratino, et al. (2020) dalam penelitiannya menggunakan metode *Support Vector Machine* dan *Naïve Bayes* untuk mengklasifikasikan komentar pada sosial media *Instagram* mengenai informasi Covid-19. Terdapat dua *software* yang digunakan dalam penelitian ini, yaitu *Instagram Scraper* untuk pengambilan data dan *software Rapid Miner* untuk melakukan analisis sentimen. Hasilnya analisis menunjukkan bahwa analisis sentimen menggunakan metode *Support Vector Machine* memiliki akurasi sebesar 80,23%, sedangkan menggunakan *Naïve Bayes* mendapatkan akurasi sebesar 78,02%.

Syarifuddin (2020) dalam penelitiannya menggunakan metode *Naïve Bayes* dan *Convolutional Neural Network (CNN)* untuk analisis sentimen opini masyarakat guna menyelaraskan dan memberi pandangan baru mengenai suatu isu tentang Covid-19, serta untuk mengetahui kecenderungan opini masyarakat di *Twitter*. Pengambilan data dilakukan dengan menggunakan *API Twitter* dan didapatkan 1098 opini dengan kata kunci “Covid-19”. Hasil penelitian ini menunjukkan bahwa metode *Naïve Bayes* memiliki akurasi dan presisi nilai positif lebih tinggi dibandingkan metode *CNN*, yaitu sebesar 63,21% dan 66,40%.

Garcia & Berton (2020) dalam penelitiannya menggunakan metode *Support Vector Machine*, *Random Forest*, dan *Logistic Regression* untuk melihat topik yang sering dibahas dan mengklasifikasikan emosi mengenai pandemi Covid-19. Data yang digunakan merupakan data dari media sosial *Twitter* yang berbahasa Inggris dan Portugal. Hasil dari penelitian ini menunjukkan bahwa terdapat sepuluh topik yang sering dibahas selama pandemi Covid-19 seperti dampak ekonomi, laporan kasus, politik, dan hiburan. Pada klasifikasi kelas

negatif ketiga metode memiliki F-1 yang sama, namun pada klasifikasi kelas positif *Linier SVM* menunjukkan performa terbaik dengan nilai *F-1 Score* sebesar 0,66 dibandingkan dengan dua metode lainnya.

Patel & Passi (2020) dalam penelitiannya membandingkan metode *Naïve Bayes*, *Support Vector Machine*, *Random Forest*, dan *K-Nearest Neighbour* untuk mengklasifikasikan ulasan mengenai Piala Dunia 2014 ke dalam tiga kelas, yaitu positif, netral, dan negatif. Data diambil dari *Twitter* pada Juni-Juli 2014 dan didapatkan sekitar 2 juta *tweet*. Hasil penelitian menunjukkan bahwa metode *Naïve Bayes* memiliki performansi yang paling baik, karena memiliki nilai akurasi dan AUC tertinggi dibandingkan metode lain.

Kajian induktif berisi tentang penelitian-penelitian terdahulu yang terkait dengan penelitian yang telah dilakukan. Berdasarkan kajian induktif yang telah dilakukan, didapatkan beberapa kelebihan terkait algoritma *Support Vector Machine* (SVM) dan *Naïve Bayes Classifier* (NBC) dalam tabel berikut:

Tabel 2. 1 Kelebihan *Support Vector Machine* dan *Naïve Bayes Classifier* pada Penelitian Sebelumnya

Penulis	Metode	Hasil Penelitian
Rofiqoh, et al., (2017)	<i>Support Vector Machine</i> dan <i>Lexicon Based Features</i>	Penelitian ini mengkombinasikan SVM dan LBF dalam mengklasifikasikan <i>tweet</i> mengenai layanan telekomunikasi. Percobaan menunjukkan bahwa metode SVM tanpa dikombinasikan dengan FBS memiliki akurasi yang lebih baik
Fanissal, et al., (2017)	<i>Naïve Bayes</i> dan <i>Query Expansion Ranking</i>	Penelitian ini mencoba mengkombinasikan metode NBC dan QER dalam mengklasifikasikan ulasan tempat pariwisata di Malang. Hasil akurasi terbaik sebesar 86,6% didapatkan pada seleksi fitur sebesar 75%.
Muthia (2017)	<i>Naïve Bayes</i> dan <i>Genetic Algorithm</i>	Penelitian menunjukkan bahwa penggunaan <i>Genetic Algorithm</i> untuk pemilih fitur berdampak pada peningkatan akurasi metode NBC pada pengklasifikasian <i>review</i> restoran.

Penulis	Metode	Hasil Penelitian
Rahmawati, et al., (2017)	<i>Support Vector Machine</i> dan <i>K-Means Clustering</i>	Mengklasifikasikan opini masyarakat mengenai pelaksanaan pilkada serentak, SVM memiliki akurasi lebih baik dari <i>K-Means Clustering</i> yaitu sebesar 91%.
Baid, et al., (2017)	<i>Naïve Bayes</i> , <i>K-Nearest Neighbour</i> , dan <i>Random Forest</i>	Mengklasifikasikan ulasan film menggunakan <i>software</i> Weka, metode <i>Naïve Bayes</i> memiliki performansi terbaik dibandingkan metode lainnya dengan akurasi sebesar 81,45%.
Permana, et al., (2017)	<i>Naïve Bayes</i>	Mengklasifikasikan <i>tweet</i> ulasan kepuasan terhadap konsultan pendidikan menunjukkan akurasi sebesar 84%.
Song, et al., (2017)	<i>Multinomial Naïve Bayes</i> (MBC), <i>Maximum Entropy</i> , dan <i>Support Vector Machine</i> (SVM)	Mengklasifikasikan 1.600.000 <i>tweet</i> ulasan novel, penelitian ini menunjukkan bahwa akurasi pengklasifikasian dipengaruhi oleh jumlah data dan pemilihan algoritma klasifikasi. Algoritma MBC menunjukkan akurasi terbaik dengan nilai sebesar 85,33%.
Haryanto, et al., (2018)	<i>Support Vector Machine</i> dan <i>Query Expansion</i>	Penelitian ini mencoba mengkombinasikan metode SVM dan QE dalam mengklasifikasikan <i>review</i> produk. Hasil percobaan menunjukkan bahwa penambahan QE berdampak pada meningkatnya akurasi metode SVM dari 94,75% menjadi 96,25%.
Al-Smadi, et al., (2018)	<i>Recurrent Neural Network</i> dan <i>Support Vector Machine</i>	Mengklasifikasikan <i>review</i> hotel di Arab, menghasilkan bahwa SVM memiliki tingkat performansi yang lebih baik dibandingkan dengan RNN berdasarkan nilai F-1 dan akurasinya.

Penulis	Metode	Hasil Penelitian
Bordoloi & Biswas (2018)	<i>Naïve Bayes</i> (NB), <i>Support Vector Machine</i> (SVM) dan <i>Maximum Entropy</i>	Mengklasifikasikan <i>review</i> produk pada tiga dataset yang berbeda, algoritma NB memiliki akurasi tertinggi pada semua dataset yang digunakan.
Alizah, et al. (2020)	<i>Naïve Bayes</i> dan <i>Support Vector Machine</i>	Hasil klasifikasi komentar pada <i>twitter</i> mengenai isu “ <i>lockdown</i> ” pada 19 April pukul 12.42-15.02 menunjukkan bahwa SVM memiliki akurasi yang lebih baik yaitu 87% daripada NB yang memiliki akurasi sebesar 81%.
Ratino, et al., (2020)	<i>Support Vector Machine</i> , <i>Naïve Bayes</i> , dan <i>Particle Swarm Optimization</i>	Penelitian ini mencoba menambahkan metode PSO pada SVM dan NB dalam mengklasifikasikan komentar mengenai informasi covid-19. Hasil penelitian menunjukkan bahwa kombinasi metode SVM dan PSO memiliki performansi terbaik yang ditunjukkan dengan nilai akurasi dan AUC tertinggi.
Syarifuddin (2020)	<i>Naïve Bayes</i> dan <i>Convolutional Neural Network</i> (CNN)	Mengklasifikasikan 1098 opini dengan kata kunci “Covid-19”, algoritma NB menunjukkan performansi lebih baik dengan nilai akurasi dan presisi nilai positif yang lebih besar daripada algoritma CNN.
Garcia & Berton (2020)	<i>Support Vector Machine</i> , <i>Random Forest</i> , dan	Berdasarkan perbandingan tiga metode tersebut diketahui bahwa nilai F-1 pada metode Linier SVM memiliki performansi yang lebih baik daripada metode yang lain dengan nilai sebesar 0,66.

Penulis	Metode	Hasil Penelitian
	<i>Logistic</i>	
	<i>Regression</i>	
Patel & Passi (2020)	<i>Naïve Bayes, Support Vector Machine, Random Forest, dan K-Neirest Neighbour</i>	Mengklasifikasikan 2 juta <i>tweet</i> mengenai <i>World Cup</i> menjadi tiga kelas yaitu positif, netral, dan negatif, algoritma NB memiliki performansi terbaik berdasarkan nilai akurasi dan AUC yang didapatkan.

Berdasarkan kajian empiris, dapat diketahui penggunaan metode *Support Vector Machine* dan *Naïve Bayes Classifier* pada penelitian terdahulu. Hasil penelitian terdahulu menunjukkan performansi dua metode tersebut yang cukup baik. Namun sampai saat ini belum ditemukan penelitian yang menggunakan metode *Support Vector Machine* dan *Naïve Bayes Classifier* untuk analisis sentimen pada *provider* layanan internet. Oleh karena itu, adanya penelitian ini diharapkan dapat menjadi gambaran dan bahan evaluasi kinerja dan pelayanan yang diberikan, khususnya layanan *IndiHome* yang ditawarkan oleh PT Telekomunikasi Indonesia.

## 2.2 Kajian Deduktif

### 2.2.1 Jasa Telekomunikasi

Secara bahasa, telekomunikasi berasal dari dua kata, yaitu “tele” yang berarti jauh dan “komunikasi” yang berarti hubungan atau pertukaran informasi (Nurdin, 2016). Dua istilah tersebut menunjukkan bahwa telekomunikasi merupakan peristiwa pertukaran informasi (komunikasi) dalam jarak yang jauh. Sedangkan secara istilah, sebagaimana yang dijelaskan pada Peraturan Menteri Komunikasi dan Informatika (Kominfo) Nomor 13 tahun 2019 tentang Penyelenggaraan Jasa. Telekomunikasi, Telekomunikasi didefinisikan sebagai setiap pemancaran, pengiriman, dan/atau penerimaan dari setiap informasi dalam bentuk tanda-tanda, isyarat, tulisan, gambar, suara, dan bunyi melalui sistem kawat, optik, radio, atau sistem elektromagnetik lainnya.

Dalam peraturan Menteri Kominfo Nomor 13 tahun 2019 tentang Penyelenggaraan Jasa, dijelaskan pula mengenai pengertian Jasa Telekomunikasi dan Penyelenggara Jaringan Telekomunikasi. Jasa Telekomunikasi didefinisikan sebagai layanan Telekomunikasi untuk memenuhi kebutuhan bertelekomunikasi dengan menggunakan Jaringan Telekomunikasi, sedangkan Penyelenggara Jaringan Telekomunikasi didefinisikan sebagai kegiatan penyedia dan/atau pelayanan Jaringan Telekomunikasi yang memungkinkan terselenggaranya Telekomunikasi.

Perkembangan teknologi informasi dan komunikasi saat ini sudah berkembang pesat. Perkembangan tersebut berdampak pada semakin ketatnya persaingan dari berbagai perusahaan penyelenggara jaringan telekomunikasi, seperti PT. Telekomunikasi Indonesia, PT Indosat Ooredoo (Indosat), PT XL Axiata Tbk (XL), PT Smartfren Telecom Tbk (Smartfren) dan sebagainya.

### 2.2.2 PT Telekomunikasi Indonesia (Telkom)

PT Telekomunikasi Indonesia, Tbk (Telkom), merupakan BUMN yang bergerak di bidang jasa pelayanan telekomunikasi dan jaringan di wilayah Indonesia. Sebagai perusahaan milik negara yang saham mayoritas perusahaan adalah Pemerintah Republik Indonesia sedangkan sisanya dikuasai oleh publik. PT Telkom juga merupakan perusahaan penyelenggara bisnis TIMES (*Telecommunication, Information, Media, Edutainment, and Service*) milik negara terbesar di Indonesia yang merupakan sebuah portofolio bisnis yang lebih lengkap mengikuti tren perubahan bisnis global di masa yang akan datang.



Gambar 2. 1 Logo PT Telkom  
(Sumber: PT Telekomunikasi Indonesia, 2019)

Sebelum menjadi PT Telkom, terdapat beberapa sejarah perusahaan sampai menjadi seperti sekarang. Berikut merupakan sejarah singkat PT Telkom:

Tabel 2. 2 Sejarah PT Telkom

Tahun	Keterangan
1856	Dimulainya pengoperasian layanan jasa telegraf elektromagnetik pertama yang menghubungkan Jakarta (Batavia) dengan Bogor (Buitenzorg).
1882	Didirikannya badan usaha swasta penyedia layanan pos dan telegraf
1961	Status jawatan diubah menjadi Perusahaan Negara Pos dan Telekomunikasi (PN Postel)
1965	PN Postel dipecah menjadi Perusahaan Negara Pos dan Giro (PN Pos & Giro) dan Perusahaan Negara Telekomunikasi (PN Telekomunikasi)
1974	PN Telekomunikasi diubah namanya menjadi Perusahaan Umum Telekomunikasi (Perumtel) yang menyelenggarakan jasa telekomunikasi nasional maupun internasional
1980	Seluruh saham PT. Indonesia Satellite Corporation Tbk. (Indosat) diambil alih oleh pemerintah RI menjadi BUMN
1989	ditetapkan Undang-undang Nomor 3 Tahun 1989 tentang Telekomunikasi, yang juga mengatur peran swasta dalam penyelenggaraan telekomunikasi
1995	Telkom menjadi perusahaan Tbk. (Terbuka)
2001	Telkom membeli 35% saham Telkomsel dari PT Indosat sebagai bagian dari implementasi restrukturisasi industri jasa telekomunikasi di Indonesia
2002	terjadi duopoli penyelenggaraan telekomunikasi lokal.

Untuk dapat menciptakan dan meningkatkan nilai bagi pelanggan, Telkom mengatur bisnis dengan berbasis pada segmen pelanggan, atau *Customer Facing Unit* (CFU). Dengan demikian, Telkom mengkategorikan portofolio produk menjadi lima segmen: *mobile*, *consumer*, *enterprise*, *wholesale international business*, dan segmen “lainnya”.



Gambar 2. 2 Portofolio Produk dan Jasa PT Telkom  
(Sumber: PT Telekomunikasi Indonesia, 2019)

PT Telkom saat ini menjadi salah satu Perusahaan Telekomunikasi terbesar di Indonesia, hal tersebut ditunjukkan dengan kepemilikannya terhadap 46 Entitas Anak. Saat ini, PT Telkom memiliki 535 outlet plasa dan 60 Witel (Wilayah Telekomunikasi) yang dibagi menjadi pada 7 divisi regional.

### 2.2.3 *Google Play*

*Google Play* merupakan layanan konten digital milik Google yang melingkupi toko online untuk produk-produk seperti musik/lagu buku, aplikasi, permainan, ataupun pemutar media berbasis *cloud*. Layanan ini dapat diakses melalui web, aplikasi android (*Play Store*) dan Google TV. *Google Play* pertama kali dirilis pada 22 Oktober 2008 dengan nama *Android Market*. Perubahan nama menjadi *Google Play* dilakukan pada Maret 2012 untuk menggantikan *Android Market* dan Layanan *Google Music*. Saat ini, jumlah aplikasi di *Google Play* mencapai 3,04 juta pada September 2020 (Clement, 2020).

Dalam *Google Play* dilengkapi fitur yang memungkinkan pengguna aplikasi dapat memberikan ulasan mengenai aplikasi yang digunakan. Ulasan ini sering digunakan sebagai alat yang efektif untuk mendapatkan informasi terhadap suatu produk atau jasa.

### 2.2.4 *Online Consumer Review*

*Online Consumer Review* merupakan salah satu bentuk dari *electronic Word of Mouth* (eWOM) dan dapat dikatakan sebagai komunikasi pemasaran yang baru dan dapat mempengaruhi serta memegang peran dalam proses keputusan pembelian, dimana informasi yang didapatkan dapat digunakan untuk mengidentifikasi dan mengevaluasi produk yang konsumen butuhkan (Sutanto & Aprianingsih, 2016). Selain itu, *online consumer review* juga didefinisikan sebagai salah satu media untuk konsumen melihat *review* dari konsumen lain terhadap suatu produk, layanan perusahaan dan tentang bagaimana sebuah perusahaan produsen (Khammash, 2008).

Menurut Ananda & Wandebori (2016) terdapat beberapa variabel dari konsep *Online Consumer Review*, yaitu:

a. *Attractiveness*

*Attractiveness* merupakan hal yang berhubungan dengan nilai sosial dari seseorang yang meliputi kepribadian, tampilan fisik, dan memiliki keadaan atau kondisi yang sama dengan *viewers*.

b. *Trustworthiness*

*Trustworthiness* adalah hal-hal yang berkaitan dengan kejujuran, integritas, dan bagaimana seorang *reviewer* dapat dipercaya.

c. *Expertise*

*Expertise* merupakan hal yang mengacu pada pengalaman, pengetahuan, dan keahlian seorang *reviewer* dalam membuktikan keunggulan dan kekurangan dari sebuah produk maupun memberikan informasi produk secara keseluruhan.

### 2.2.5 *Web Scraping*

*Web Scraping* adalah proses pengambilan sebuah dokumen semi-terstruktur dari internet, umumnya berupa halaman-halaman *website* dalam bahasa *markup* seperti HTML (*HyperText Markup Language*) atau XHTML (*Extensible HyperText Markup Language*), dan menganalisa dokumen tersebut untuk diambil data tertentu dari halaman tersebut untuk digunakan bagi kepentingan lain (Turland, 2010). Bahasa *markup* merupakan bahasa komputer yang menggunakan *tag* untuk mendefinisikan elemen dalam dokumen. (Josi, et al., 2014)

Menurut Josi, et al. (2014) terdapat beberapa langkah dalam proses *web scraping*, yaitu:

1. *Create Scraping Template*

Pembuat program mempelajari dokumen HTML dari *website* yang akan diambil informasinya dari *tag* HTML yang mengapit informasi yang akan diambil.

2. *Explore Site Navigation*

Pembuat program mempelajari teknis navigasi pada *website* yang akan diambil informasinya untuk ditirukan pada aplikasi *web scraper*.

3. *Automate Navigation and Extraction*

Berdasarkan informasi yang didapatkan dari langkah 1 dan 2 diatas, aplikasi *web scraper* dibuat untuk mengotomatisasi pengambilan informasi dari *website* yang ditentukan.

4. *Extracted Data and Package History*

Informasi yang didapat dari langkah 3 disimpan dalam tabel atau tabel-tabel *database*.

### 2.2.6 *Machine Learning*

*Machine Learning* adalah ilmu yang mempelajari bagaimana komputer dapat belajar atau meningkatkan kinerja berdasarkan data agar secara otomatis mengenali pola data untuk membuat keputusan cerdas berdasarkan data (Han & Kember, 2012). *Machine Learning* merupakan suatu area dalam *Artificial Intelligence* (AI) yang berhubungan dengan pengembangan teknik-teknik yang bisa diprogramkan dan belajar dari data masa lalu (Santoso, 2007)

Menurut Mohri, et al. (2012) *machine learning* dapat didefinisikan sebagai metode komputasi berdasarkan pengalaman untuk meningkatkan performa atau membuat prediksi yang akurat. Pengalaman disini didefinisikan sebagai data historis yang tersedia dan dapat dijadikan sebagai data pembelajaran (*training data*). Dalam pembelajaran *machine learning*, terdapat beberapa skenario-skenario, seperti:

a. *Supervised Learning*

Penggunaan skenario *supervised learning*, pembelajaran menggunakan masukan data pembelajaran yang telah diberi label. Setelah itu membuat prediksi dari data yang telah diberi label.

b. *Unsupervised Learning*

Penggunaan skenario *unsupervised learning*, pembelajaran menggunakan masukan data pembelajaran yang tidak diberi label. Setelah itu mencoba untuk mengelompokkan data berdasarkan karakteristik-karakteristik yang ditemui.

c. *Reinforcement Learning*

Pada skenario *reinforcement learning* fase pembelajaran dan tes saling dicampur. Untuk mengumpulkan informasi pembelajaran secara aktif dengan berinteraksi ke lingkungan sehingga untuk mendapatkan balasan untuk setiap aksi dari pembelajaran.

### 2.2.7 *Text Mining*

*Text Mining* dapat diartikan sebagai penemuan informasi baru yang sebelumnya tidak diketahui oleh komputer dengan mengekstrak informasi secara otomatis dari sumber yang berbeda. Kunci dari proses ini adalah menggabungkan informasi yang berhasil diekstraksi dari berbagai sumber (Hearst, 2003). Menurut Harlian (2006) *text mining* didefinisikan sebagai data berupa teks yang

biasanya sumber data didapatkan dari dokumen, dengan tujuan adalah mencari kata-kata yang dapat mewakili isi dari dokumen tersebut yang nantinya dapat dilakukan analisa hubungan antar dokumen.

Dalam proses *text mining*, teks dokumen yang akan digunakan harus dipersiapkan terlebih dahulu sebelum dapat digunakan untuk proses utama. Proses mempersiapkan teks dokumen atau *dataset* mentah disebut juga dengan proses *text preprocessing*. *Text preprocessing* berfungsi untuk mengubah data teks yang tidak terstruktur atau sembarangan menjadi data yang terstruktur. Secara umum proses yang dilakukan dalam tahapan preprocessing adalah sebagai berikut:

1. *Spelling Normalization*

*Spelling Normalization* adalah proses substitusi atau perbaikan kata-kata singkatan atau salah ejaan. Substitusi kata dilakukan untuk menghindari jumlah perhitungan dimensi kata yang melebar. Perhitungan dimensi kata akan melebar jika kata yang salah eja atau disingkat tidak diubah karena kata tersebut sebenarnya memiliki maksud dan arti yang sama tetapi akan dianggap sebagai entitas yang berbeda pada saat proses penyusunan matriks.

2. *Case Folding*

*Case Folding* merupakan proses penyamaan case dalam sebuah dokumen. Hal ini dilakukan untuk mempermudah pencarian. Tidak semua dokumen teks konsisten dalam penggunaan huruf kapital. Oleh karena itu peran *case folding* dibutuhkan dalam mengkonversi keseluruhan teks dalam dokumen menjadi suatu bentuk standar, dalam hal ini huruf kecil atau *lowercase*.

3. *Tokenizing*

*Tokenizing* adalah proses penguraian teks yang semula berupa kalimat-kalimat yang berisi kata-kata. Proses ini diawali dengan menghilangkan delimiter-delimiter yaitu simbol dan tanda baca yang ada pada teks tersebut seperti @, \$, &, tanda titik (.), koma (,), tanda tanya (?), tanda seru (!). Proses pemotongan *string* berdasarkan tiap kata yang menyusunnya, pada umumnya setiap kata akan terpisahkan dengan karakter spasi, proses tekonisasi mengandalkan karakter spasi pada dokumen teks untuk melakukan pemisahan. Hasil dari proses ini adalah kumpulan kata saja (Putri, 2016).

#### 4. *Filtering*

*Filtering* adalah proses mengambil kata-kata penting dari hasil token. Algoritma *stoplist/stopword* (membuang kata yang kurang penting) atau *wordlist* (menyimpan kata yang penting) dapat digunakan pada tahap ini. *Stopword* adalah kata-kata yang tidak deskriptif dan bukan merupakan kata penting dari suatu dokumen sehingga dapat dibuang. Contoh *stopword* adalah “yang”, “dan”, “di”, “dari”, dan lain sebagainya (Putri, 2016). Dalam proses ini menggunakan *stopword* agar kata-kata yang kurang penting dan sering muncul dalam suatu dokumen dibuang sehingga hanya menyisakan kata-kata yang penting dan mempunyai arti yang diproses ke tahap selanjutnya.

#### 5. *Stemming*

*Stemming* merupakan proses pengolahan kata untuk mendapatkan kata dasar dari sebuah kata yang telah mengalami imbuhan dengan asumsi bahwa kata-kata tersebut sebenarnya memiliki makna dan arti yang sama. Algoritma ini bekerja berdasarkan struktural morfologi dalam kalimat Bahasa Indonesia, yang terdiri dari awalan, akhiran, sisipan, dan awalan+akhiran. Tujuan dari tahap ini adalah:

- a. Keefisiensian, pada *stemming* dilakukan pengurangan jumlah kata dalam dokumen agar mengurangi kebutuhan dalam ruang penyimpanan dan mempercepat dalam melakukan pencarian.
- b. Keefektifan, *stemming* dilakukan untuk mengurangi *recall* dengan pengurangan bentuk-bentuk kata ke dalam bentuk dasarnya (Putri, 2016). Sebagai contoh adalah kata “duduk-lah”, “minum-lah”, “jika-pun”, dan sebagainya.

#### 2.2.8 *Sentiment Analysis*

Menurut Pang & Lee (2008), analisis sentimen atau *opinion mining* merupakan proses memperoleh informasi dengan cara memahami, mengekstrak, dan mengolah data tekstual secara otomatis. Analisis sentimen mulai terkenal pada tahun 2013 sebagai salah satu cabang *text mining*. Pada dasarnya analisis sentimen digunakan untuk mengetahui tanggapan dan sikap dari suatu kelompok atau individu terhadap suatu topik bahasan kontekstual keseluruhan dokumen. Tanggapan dan sikap tersebut dapat berupa pendapat, penilaian, evaluasi,

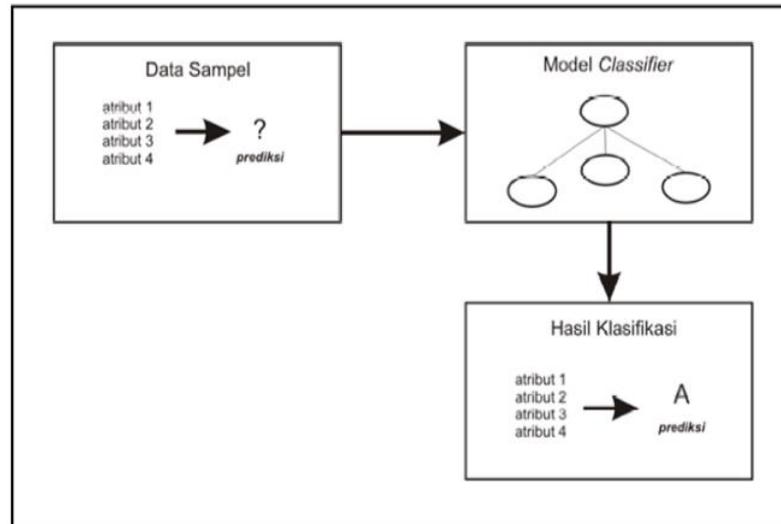
keadaan afektif (keadaan emosional penulis saat menulis), atau komunikasi emosional (efek emosional yang sampai pada pembaca) (Saraswati, 2011).

Bagi sektor bisnis, analisis sentimen dapat berguna dalam proses pelacakan produk, jasa, merek, dan target konsumen di pasar. Selain itu, analisis sentimen juga dapat menilai keunggulan dan kelemahan suatu produk dan jasa. Secara umum, analisis sentimen digunakan untuk mendeteksi keluhan, persepsi produk atau layanan baru, dan persepsi dari suatu merek tertentu.

### 2.2.9 Klasifikasi

Menurut Prasetyo (2012), klasifikasi adalah proses pengelompokan teramati dari suatu objek data ke dalam suatu kelas tertentu berdasarkan kelas-kelas yang ada. Sedangkan menurut Han & Kamber (2006) klasifikasi didefinisikan sebagai teknik analisis data yang digunakan untuk menghasilkan model prediksi untuk mendeskripsikan label atau kelas data.

Menurut Han & Kamber (2006) terdapat dua tahap dalam proses klasifikasi. Tahap pertama merupakan *learned model*, pada tahap ini dilakukan pembangunan model menggunakan hasil analisa *record database* dari serangkaian kelas yang ada. Masing-masing *record* diasumsikan mempunyai *predefined class* yang didasarkan pada atribut kelas label, oleh karena itu klasifikasi termasuk dalam *supervised learning*. Tahap ini juga disebut sebagai tahap pembelajaran atau pelatihan. Sebuah algoritma klasifikasi akan membangun sebuah model klasifikasi dengan cara menganalisis data *training*. Tahap pembelajaran dapat juga dipandang sebagai tahap pembentukan fungsi atau pemetaan  $y = f(x)$ , dimana  $y$  merupakan kelas hasil prediksi dan  $x$  adalah *record* yang ingin diprediksi kelasnya.



Gambar 2. 3 Proses Klasifikasi  
(Sumber: Han & Kamber, 2006)

Kualitas hasil kualifikasi dapat dinilai dan dievaluasi berdasarkan beberapa ukuran:

1. *Accuracy*

Akurasi adalah jumlah proporsi prediksi yang benar. Akurasi digunakan sebagai tingkat ketepatan antara nilai aktual dengan nilai prediksi.

2. *Precision*

Presisi adalah proporsi jumlah dokumen teks yang relevan terkendali diantara semua dokumen yang dipilih sistem. Presisi digunakan sebagai tingkat ketepatan anata informasi yang diminta dengan jawaban yang diberikan oleh sistem (Manning, et al., 2009).

3. *Recall*

*Recall* adalah proporsi jumlah dokumen teks yang relevan terkendali diantara semua dokumen teks relevan yang ada pada koleksi. *Recall* digunakan sebagai ukuran keberhasilan sistem dalam menemukan kembali informasi (Manning, et al., 2009).

4. *F-Measure*

*F-Measure* adalah merupakan kata-kata harmonis dari nilai *recall* dan nilai presisi sehingga dapat memberikan penilaian kinerja yang lebih seimbang. *F-Measure* digunakan untuk mengukur kinerja sistem secara menyeluruh dalam pengklasifikasian (Manning, et al., 2009).

### 2.2.10 Pembobotan

Pembobotan merupakan metode untuk mengubah input data menjadi suatu fitur vektor. Metode pembobotan yang umum digunakan adalah *bag-of-feature*. Sebagai contoh terdapat sederet fitur seperti pada vektor  $\{f_1, f_2, \dots, f_n\}$  dimana yaitu sekumpulan fitur-fitur sebanyak  $n$  yang telah ditentukan sebelumnya. Misalkan kata “puas” maka fitur vektor dari data adalah vektor. Terdapat beberapa metode pembobotan, seperti:

a. *Term Presence* (TP)

*Term Presence* (TP) merupakan metode pembobotan pada suatu dokumen teks yang melihat keberadaan daftar kata-kata (term) atau fitur yang ada pada corpus terhadap suatu dokumen. Jika fitur yang ada pada daftar fitur acuan terdapat pada dokumen yang sedang diboboti maka nilai fitur tersebut pada *feature vector* akan diberi nilai 1 dan tidak menghiraukan jumlah kemunculan fitur tersebut. Jika fitur tersebut tidak ada pada dokumen maka nilai 0 pada *feature space* (O’Keefe & Koprinska, 2009).

b. *Term Frequency* (TF)

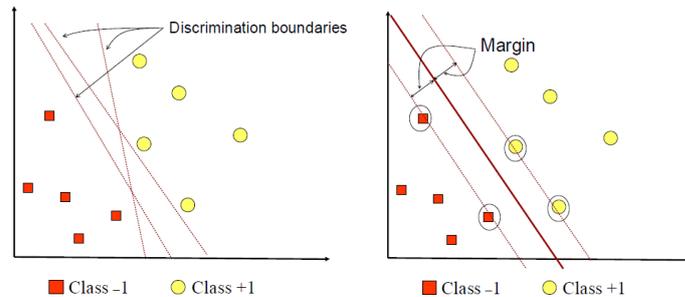
*Term Frequency* (TF) memiliki kesamaan dengan TP yang telah dijelaskan sebelumnya, tapi yang membedakan adalah TF menghitung jumlah kemunculan fitur acuan pada suatu dokumen bukan hanya keberadaan fitur tersebut (O’Keefe & Koprinska, 2009).

### 2.2.11 Support Vector Machine

*Support Vector Machine* (SVM) adalah teknik yang relatif baru untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi (Handayani, et al., 2020). SVM termasuk dalam jenis klasifikator yang biner, linier, dan non probabilistik. Pada umumnya hal pertama yang mendasari untuk memahami klasifikasi dengan SVM adalah mencari garis (*hyperplane*) yang optimal (Mutawalli, et al., 2019). Hal ini berfungsi untuk memisahkan dua kelas data yang berbeda, yaitu positif (+1) dan negatif (-1). Pada gambar 2.4 untuk data positif (+1) disimbolkan dengan warna kuning dan data negatif (-1) disimbolkan dengan warna merah.

Secara umum, proses SVM diilustrasikan dalam gambar 2.4. Grafik sebelah kiri pada gambar 2.4 mengilustrasikan tentang beberapa kemungkinan garis pemisah (*discrimination boundaries*) pada SVM untuk melakukan set data.

Sedangkan pada grafik sebelah kanan mengilustrasikan *discrimination boundaries* dengan margin maksimum. Margin atau juga yang disebut juga batas pemisah adalah jarak antara dua kelas data terdekat pada bidang *hyperplane*. *Hyperplane* dengan margin terbaik yang akan memberikan generalisasi untuk mendapatkan hasil klasifikasi yang lebih baik.



Gambar 2. 4 Proses SVM dalam Menemukan *Hyperline*  
Sumber: (Handayani, et al., 2020)

### 2.2.12 Naïve Bayes Classifier

*Naïve Bayes Classifier* (NBC) merupakan pengklasifikasian probabilitas sederhana berdasarkan pada teorema *Bayes*. Teorema *Bayes* dikombinasikan dengan “*Naive*” yang berarti setiap atribut atau variabel bersifat bebas (*independent*). NBC dapat dilatih dengan efisien dalam pembelajaran terawasi. Proses NBC mengasumsikan bahwa ada atau tidaknya suatu fitur pada suatu kelas tidak berhubungan dengan ada atau tidaknya fitur lain dikelas yang sama (Setiawan, 2014). Algoritma NBC termasuk dalam pembelajaran *supervised*, sehingga pada tahapan pembelajaran dibutuhkan data awal berupa data pelatihan untuk dapat mengambil keputusan. Pada tahap pengklasifikasian akan dihitung nilai probabilitas dari masing-masing label kelas yang ada terhadap masukan yang diberikan. Label kelas yang memiliki nilai probabilitas paling besar akan dijadikan label kelas data masukan tersebut.

*Naïve Bayes Classifier* merupakan perhitungan teorema *Bayes* yang paling sederhana, karena mampu mengurangi kompleksitas komputasi menjadi multiplikasi sederhana dari probabilitas (Arifin & Sasongko, 2018). Selain itu, algoritma ini juga mampu menangani set data yang memiliki banyak atribut . Berikut merupakan persamaan NBC (Arifin & Sasongko, 2018):

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (2.1)$$

dengan:

$X$  = Data dengan *class* yang belum diketahui

$H$  = Hipotesis data  $X$  merupakan suatu kelas spesifik

$P(H|Q)$  = Probabilitas hipotesis  $H$  berdasarkan kondisi  $x$  (posteriori prob.)

$P(H)$  = Probabilitas hipotesis  $H$  (prior prob.)

$P(X|H)$  = Probabilitas  $X$  berdasarkan kondisi tersebut

$P(X)$  = Probabilitas  $X$

### 2.2.13 Asosiasi Kata

Asosiasi kata diperoleh dengan melakukan pendekatan pada perhitungan nilai korelasi. Pada umumnya, nilai korelasi digunakan untuk menyatakan hubungan dua atau lebih variabel kuantitatif, namun pada asosiasi kata nilai korelasi dimaknai sebagai keeratan hubungan antara dua atau lebih variabel kuantitatif (Ulwan, 2016). Korelasi bertujuan untuk menemukan tingkat hubungan antara variabel bebas ( $X$ ) dan variabel bebas ( $Y$ ), dalam ketentuan data memiliki syarat-syarat tertentu (Fadliyah, 2014). Pendekatan asosiasi pada penelitian ini digunakan untuk mencari hubungan antar teks *review* aplikasi *MyIndiHome* sehingga akan didapatkan informasi yang bermanfaat bagi pihak terkait.

### 2.2.14 Diagram *Fishbone*

Diagram *Fishbone* atau sering dikenal dengan “Diagram Ishikawa” karena dikembangkan oleh Ishikawa pada sekitar tahun 1960-an. Diagram ini disebut *fishbone* karena memiliki kerangka yang menyerupai tulang ikan yang meliputi kepala, sirip dan duri. Diagram *fishbone* merupakan alat visual yang digunakan untuk mengidentifikasi, mengeksplorasi yang secara grafik menggambarkan semua penyebab suatu permasalahan. Konsep dasar dari diagram *fishbone* adalah permasalahan mendasar yang diletakkan pada bagian kanan dari diagram atau pada bagian kepala dari kerangka tulang ikannya. Sementara sirip dan duri menggambarkan sebagai sebab permasalahannya. Diagram *fishbone* umumnya digunakan untuk mengidentifikasi permasalahan dan menentukan penyebab dari permasalahan tersebut (Fritz, 2016).

## BAB III METODE PENELITIAN

### 3.1 Objek dan Subjek Penelitian

Objek penelitian ini adalah persepsi pengguna layanan *IndiHome* terhadap pelayanan *IndiHome* melalui *review* pada *database website Google Play*. Sedangkan subjek pada penelitian ini merupakan pengguna *IndiHome*.

### 3.2 Populasi dan Sampel Penelitian

Populasi dalam penelitian ini adalah semua ulasan atau *review* pengguna layanan *IndiHome* dari *database website Google Play* melalui aplikasinya yaitu *MyIndiHome*. Sedangkan untuk sampel yang digunakan adalah ulasan aplikasi *MyIndiHome* pada tanggal 1 November 2020 – 15 Desember 2020 karena waktu tersebut merupakan periode penggunaan aplikasi versi 3.85.005. Versi tersebut merupakan versi aplikasi terakhir sebelum dilakukan *update* terbaru.

### 3.3 Metode Pengumpulan Data

Pada penelitian ini, metode pengumpulan data yang digunakan adalah metode *web scraping*. Alat yang digunakan dalam metode ini *Data Scraper* yang merupakan *extension* dari *Google Chrome*.

### 3.4 Jenis dan Sumber Data

Jenis data yang digunakan dalam penelitian ini merupakan data primer. Data tersebut diperoleh dengan menggunakan metode *scraping* dari situs *website MyIndiHome* menggunakan *extensions Google Chrome* yaitu *Data Miner* pada alamat website

<https://play.google.com/store/apps/details?id=com.telkom.indihome.external&showAllReviews=true>. Data yang diperoleh berupa *review* atau ulasan pengguna aplikasi *MyIndiHome* yang diambil dari tanggal 1 November 2020 – 15 Desember 2020 sebanyak 2.539 ulasan.

### 3.5 Variabel Penelitian

Pada penelitian ini terdapat dua macam variabel yang digunakan, yaitu:

- a. *Date*, merupakan tanggal dibuatnya ulasan
- b. *Review*, merupakan isi ulasan pengguna.

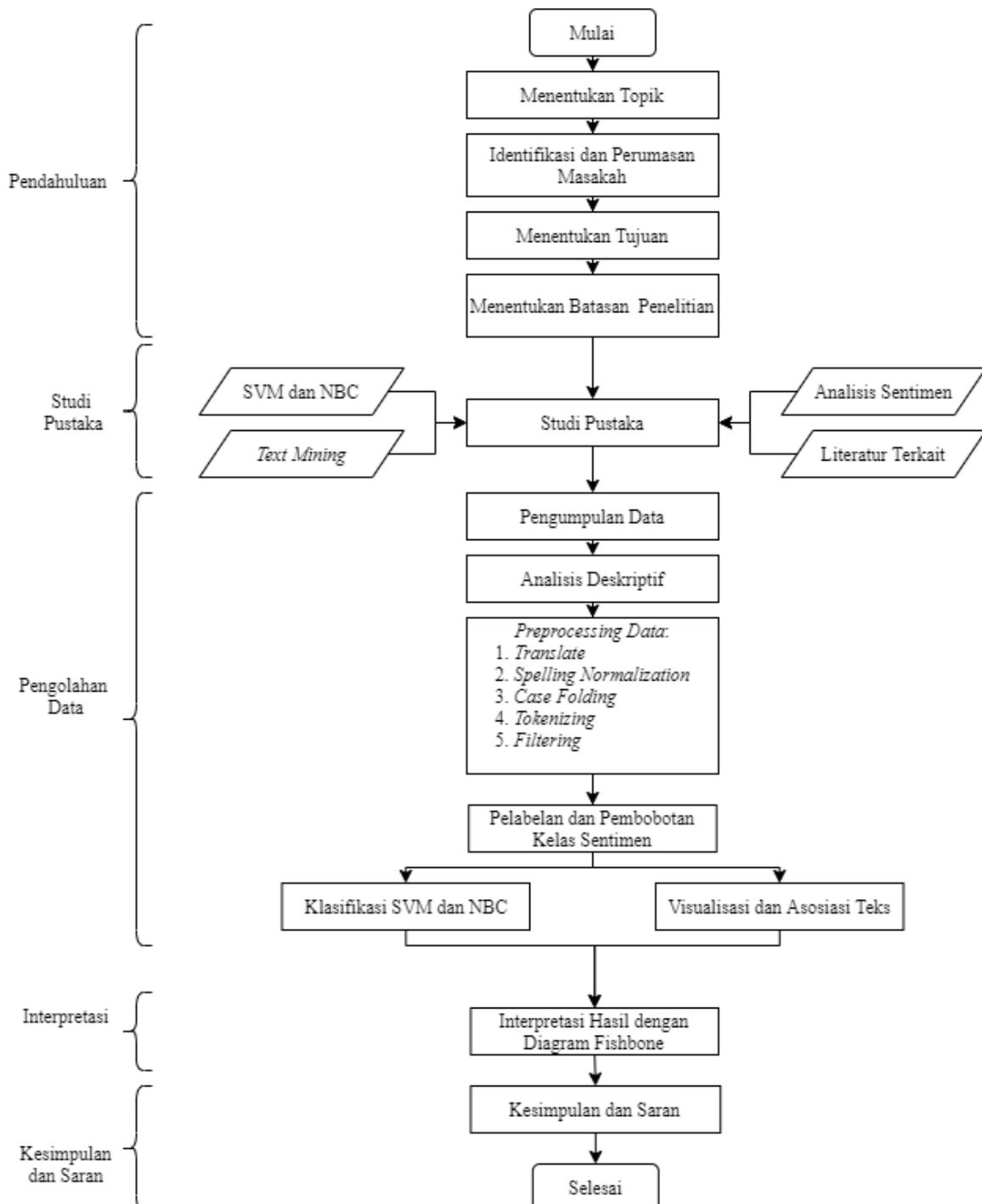
### 3.6 Metode Analisis Data

Dalam penelitian ini, digunakan software R Studio versi 1.4.1103 dan *Microsoft Excel* 2016. Terdapat beberapa metode analisis data yang digunakan dalam penelitian ini, antara lain:

- a. Analisis deskriptif, digunakan untuk memberikan gambaran ulasan *MyIndiHome* yang ada pada situs *Google Play*.
- b. Analisis sentimen, digunakan untuk melakukan pelabelan data ke dalam kelas sentimen positif dan negatif.
- c. Metode *machine learning*, yaitu metode *Support Vector Machine* (SVM) dan *Naïve Bayes Classifier* (NBC) digunakan untuk mengklasifikasikan ulasan yang berbentuk positif dan negatif.
- d. *Wordcloud*, digunakan untuk melakukan visualisasi kata yang paling sering muncul digunakan dalam ulasan.
- e. Diagram *Fishbone*, digunakan untuk mengidentifikasi faktor-faktor penyebab permasalahan yang didapatkan dari ulasan negatif sehingga dapat dilakukan pemecahan masalah yang dihadapi.

### 3.7 Tahap Penelitian

*Flowchart* yang menunjukkan tahapan penelitian dari awal sampai akhir ditunjukkan oleh Gambar 3.1.



Gambar 3. 1 *Flowchart* Penelitian

Berdasarkan *flowchart* penelitian, berikut merupakan langkah-langkah pada penelitian ini:

1. Mulai

## 2. Menentukan Topik

Proses penelitian ini diawali dengan menentukan topik penelitian yang akan dilakukan. Topik penelitian ini yaitu analisis sentimen pada layanan jasa internet yang diberikan oleh perusahaan BUMN di bidang telekomunikasi, yaitu layanan *IndiHome* dari PT Telekomunikasi Indonesia Tbk.

## 3. Identifikasi dan Perumusan Masalah

Pada tahap ini dilakukan identifikasi masalah dengan melihat kondisi dan situasi saat. Hasil identifikasi didapatkan informasi bahwa ketika banyak perusahaan yang mengalami penurunan *income* ketika terjadi pandemi, layanan *IndiHome* dari PT Telekomunikasi Indonesia menjadi salah satu yang mengalami kenaikan *income* yang cukup drastis. Namun, berdasarkan rating pada aplikasi *myIndiHome* menunjukkan layanan yang diberikan belum sesuai harapan *customer*. Selanjutnya dilakukan perumusan masalah yang akan menjadi pedoman dalam penelitian ini. Berdasarkan permasalahan yang didapatkan, terdapat beberapa rumusan masalah yang terbentuk, seperti bagaimana gambaran umum mengenai persepsi pengguna layanan *IndiHome* berdasarkan aplikasi *myIndiHome* pada *website Google Play*, bagaimana performa metode SVM dan NBC dalam mengklasifikasikan data ulasan pengguna layanan *IndiHome* menjadi kelas positif dan negatif, informasi apa yang didapatkan dalam setiap klasifikasi yang telah dilakukan, faktor apa saja yang perlu diperbaiki berdasarkan hasil ulasan negatif yang didapat?

## 4. Menentukan Tujuan

Pada tahap ini dilakukan penentuan tujuan penelitian. Tujuan penelitian ini menjelaskan hal-hal yang ingin dicapai dalam penelitian ini. Adapun tujuan dalam penelitian ini adalah mengetahui gambaran umum mengenai persepsi pengguna layanan *IndiHome* berdasarkan aplikasi *myIndiHome* pada *website Google Play*, mengetahui performa metode SVM dan NBC dalam mengklasifikasikan data ulasan pengguna layanan *IndiHome* menjadi kelas positif dan negatif, mendapatkan informasi yang bermanfaat dalam setiap klasifikasi yang telah dilakukan, dan mengetahui faktor-faktor yang perlu diperbaiki berdasarkan hasil ulasan yang didapat

#### 5. Menentukan Batasan Penelitian

Dalam penelitian ini digunakan batasan penelitian agar penelitian tetap terarah dan pembahasan tidak melebar kepada hal yang tidak perlu. Adapun batasan dalam penelitian ini yaitu data yang digunakan merupakan data ulasan mengenai aplikasi *myIndiHome* pada situs *Google Play* yang diambil pada tanggal 1 November 2020 – 15 Desember 2020 dan ulasan yang diambil merupakan ulasan yang menggunakan bahasa Indonesia

#### 6. Studi Pustaka

Pada tahap ini dilakukan studi pustaka yang berkaitan dengan topik analisis sentimen untuk dapat mengetahui metode yang sesuai untuk penelitian ini. Terdapat beberapa topik yang digunakan dalam studi pustaka, seperti *Text Mining*, *Google Play*, dan *Wordcloud*.

#### 7. Pengumpulan Data

Pada tahap ini dilakukan pengumpulan data dengan melakukan ekstraksi ulasan secara online dengan teknik *scraping data* dengan bantuan ekstensi dari *Google Chrome* yaitu *Data Scraper*. Data yang digunakan yaitu berupa ulasan pada aplikasi *myIndiHome* pada *website Google Play*.

#### 8. Analisis Deskriptif

Analisis Deskriptif digunakan untuk menggambarkan ulasan secara umum dari ulasan pengguna *IndiHome* seperti jumlah *review* di setiap periode.

#### 9. *Preprocessing Data*

*Preprocessing data* merupakan tahap awal dalam proses pengolahan data. Proses ini dilakukan untuk menyeleksi data dan mengubahnya menjadi data yang lebih terstruktur dengan bantuan *software R*. Terdapat beberapa proses yang dilakukan dalam tahap ini, seperti:

- a. *Translate* : tahap mengubah ulasan yang berbahasa Inggris maupun campuran (Inggris dan Indonesia) akan diubah menjadi bahasa Indonesia.
- b. *Spelling Normalization* : memperbaiki kata-kata yang disingkat maupun salah ejaan.
- c. *Case Folding* : tahap mengubah teks menjadi *lowercase*, menghapus angka, dan menghapus tanda baca.
- d. *Tokenizing* : proses pemisahan teks ulasan menjadi kata per kata yang tidak saling berpengaruh (*independent*)

e. *Filtering* : proses penyaringan atau pemilihan kata dalam dokumen

#### 10. Pelabelan dan Pembobotan Kelas Sentimen

Tahap ini diawali dengan menghitung jumlah skor setiap ulasan berdasarkan kamus *lexicon*. Setelah itu akan dilakukan pelabelan kelas dengan ketentuan, apabila skor ulasan kurang dari 0 maka termasuk kelas negatif, sama dengan 0 maka termasuk kelas netral, dan lebih dari 0 maka termasuk kelas positif.

#### 11. Klasifikasi SVM dan NBC

Data yang telah memiliki label kelas, akan digunakan untuk proses klasifikasi. Proses ini dilakukan untuk membentuk model yang akan digunakan untuk mengklasifikasikan data baru. Proses ini dilakukan dengan dua metode, yaitu *Support Vector Machine (SVM)* dan *Naïve Bayes Classifier (NBC)* dan dilakukan dengan bantuan *software R*.

#### 12. Visualisasi dan Asosiasi Teks

Tahap ini dilakukan dengan mengekstraksi seluruh informasi sehingga dapat diketahui kata yang paling sering dibahas pada ulasan pengguna. *Output* dari tahap ini nantinya berupa *wordcloud* dan *barplot*.

#### 13. Interpretasi Hasil dengan Diagram *Fishbone*

Setelah dilakukan tahap visualisasi dan asosiasi teks pada kelas sentimen negatif, akan didapatkan informasi mengenai topik bahasan pengguna yang menjadikan pengguna memberikan ulasan negatif. Berdasarkan informasi tersebut, akan dilakukan analisis faktor-faktor yang menjadi masalah menurut pengguna *IndiHome* menggunakan diagram *fishbone* dan mengidentifikasi usulan perbaikan yang perlu dilakukan perusahaan untuk menangani masalah tersebut.

#### 14. Kesimpulan dan Saran

Penelitian ini diakhiri dengan tahap membuat kesimpulan dan saran berdasarkan hasil penelitian.

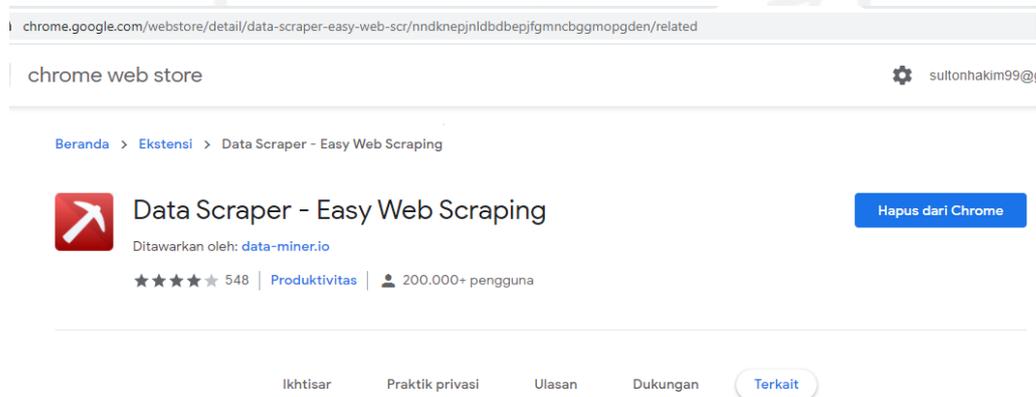
#### 15. Selesai

## BAB IV

### PENGUMPULAN DAN PENGOLAHAN DATA

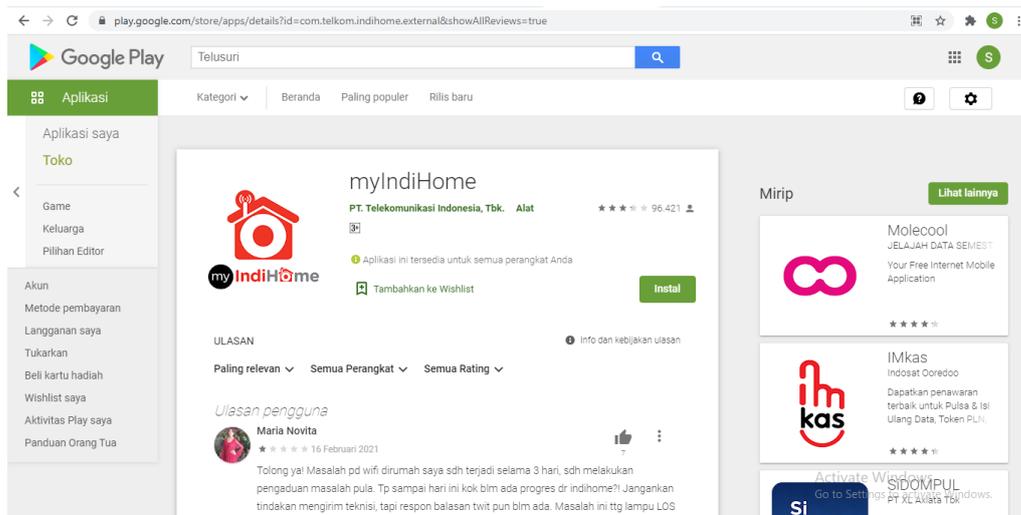
#### 4.1 Pengumpulan Data

Data berupa ulasan pengguna layanan IndiHome melalui aplikasi myIndiHome diambil pada situs *Google Play*. Pengambilan data dilakukan menggunakan teknik *scraping* dengan memanfaatkan ekstensi dari *Google Chrome*. Terdapat banyak ekstensi dari *Google Chrome* yang dapat digunakan untuk mengambil data pada situs online secara gratis, salah satunya yaitu *Data Scraper*. Pada penelitian ini digunakan *Data Scraper* versi 5.2.74 untuk mengambil data ulasan yang selanjutnya diimpor menjadi *spreadsheet Microsoft Excel* atau CSV.



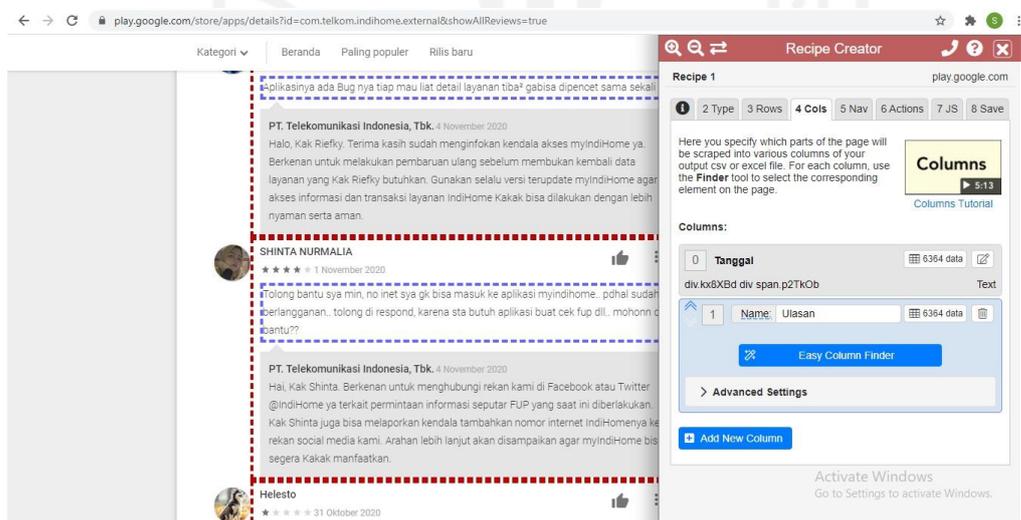
Gambar 4. 1 Ekstensi *Data Scraper*

Langkah awal yang dilakukan adalah menambahkan atau mengaktifkan ekstensi *Data Scraper* pada *Google Chrome*. Kemudian dapat dilakukan pengambilan data sesuai alamat *domain* yang diinginkan. Pada penelitian ini, data yang digunakan diambil dari ulasan pengguna IndiHome dengan alamat domain <https://play.google.com/store/apps/details?id=com.telkom.indihome.external&showAllReviews=true>.



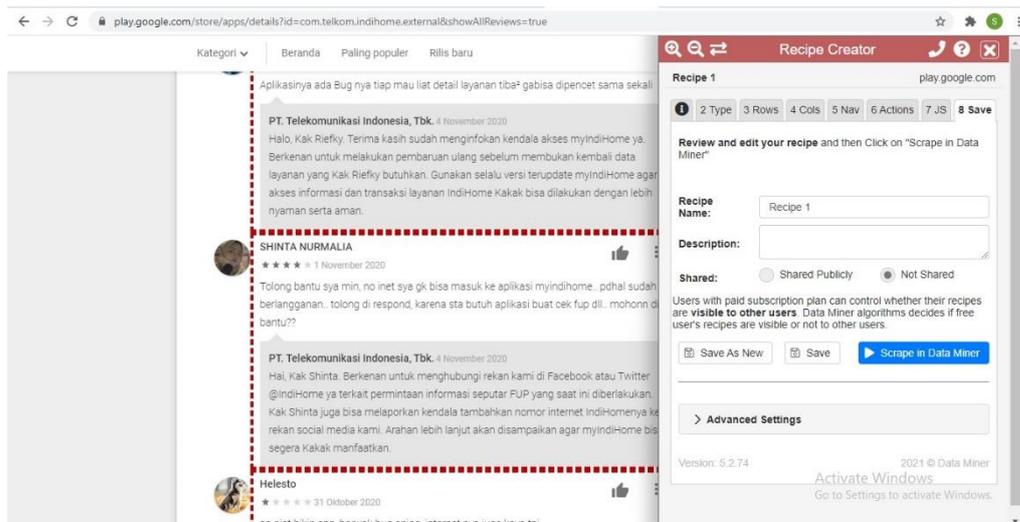
Gambar 4. 2 Website myIndiHome

Langkah selanjutnya adalah membuka alamat domain atau website seperti gambar 4.2 kemudian memilih data ulasan sesuai dengan batasan yang telah ditentukan. Pertama pada *start* memilih sesuai dengan data pada *website*. Selanjutnya pada menu *Cols* memilih sesuai dengan data yang akan diambil, dalam penelitian ini data yang diambil berupa data waktu (tanggal) dan komentar. Pembuatan setiap *query* pada *Cols* akan terdapat tampilan seperti Gambar 4.3.



Gambar 4. 3 Pembuatan *Query*

Pada penelitian ini terdapat dua kolom, yaitu tanggal dan ulasan. Selanjutnya dapat dilakukan *scraping* dengan memilih menu *Save*, mengisi *Name*, dan terakhir klik *Scrape in Data Miner*.

Gambar 4. 4 Proses *Srcaping*

Hasil data yang telah diekstraksi dapat disimpan dalam bentuk xlsx atau csv. Pada penelitian ini data disimpan dalam bentuk xlsx.

1	Tanggal	Ulasan
2	15 Desember 2020	Habis update kok malah eror mulu, dari aplikasi ngfreeze sampai eror login, padahal sudah sesuai nomor hp. Tolong perbaiki masa ga bs di pencet aplikasinya, jani
3	15 Desember 2020	Taaaik,bayar mahal"ngeleek yg adaa
4	15 Desember 2020	Kok nmr Indihome sy katanya TDK terdaftar padahal nmr yg sy masukkan sdn benennrrr
5	15 Desember 2020	tolong dperbaiki pelayanannya, masa malem" dbuat nge game pake wifi lemot parah bayar doank yg dlihat, tp kenyamanan pelanggan gk dperhatikan
6	15 Desember 2020	Internet gk sampe 1 MB hedeheh
7	15 Desember 2020	Sampah lambat bet indihome sekarang
8	15 Desember 2020	Apaan dikit ngedown lambat banget Pake Layanan 10 mbs tpi kecepatannya aj gk nyampe 1 MB
9	15 Desember 2020	apk kalo gak mau maju ya gini.....mau masang nomor indihome nya j g bisa .....tidak ad cara untuk mempermudah pemakai apk adalah awal dari kepergian sipemal
10	15 Desember 2020	Aplikasinya produk gagal
11	15 Desember 2020	lumayan lah,tapi tidak seperti dulu cek batas dan pemakaian ada...tapi sekarang cuman bisa liat pemakaian,jadi tidak tau batasnya kuota.mohon di benahi lagi bi
12	15 Desember 2020	Payah..! Setiap mau laporkan gangguan selalu muncul kalimat Maaf layanan indihome anda terisolir pastikan tdk ada pembayaran yg tertunda, Padahal tiap bulan
13	15 Desember 2020	Tidak bisa cek kuota pengguna internet IndiHome saya di appl.. Mohon di bantu
14	15 Desember 2020	Mantapp
15	15 Desember 2020	Cepat pelayanan y .....
16	15 Desember 2020	Mengecek penggunaan kadang gak jalan
17	15 Desember 2020	Indihome parah bgt saya sudah bayar kemaren sebesar 289.000sampai saat nih tdk bs digunakan untuk internet.. Tolong diperbaiki..
18	15 Desember 2020	Ok
19	15 Desember 2020	Sangat mencecewakan. kalo ada bintang minus itu lebih pantas

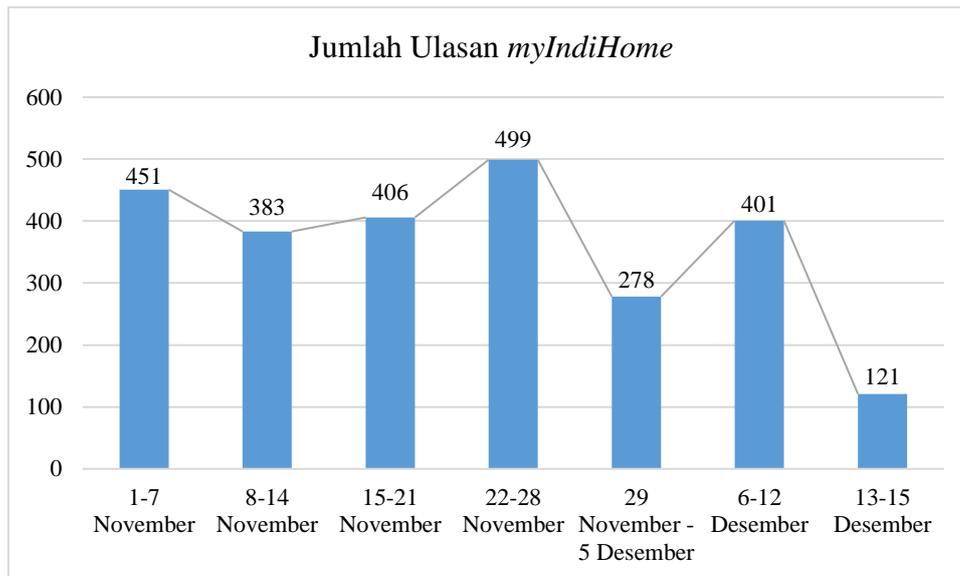
Gambar 4. 5 Hasil *Scraping* Data

## 4.2 Pengumpulan Data

Pengolahan data dilakukan dengan melakukan analisis deskriptif sebanyak 2.539 ulasan *myIndiHome* pada situs *Google Play*. Kemudian, data yang telah didapatkan akan dilakukan *pre-processing* data yang meliputi *translate*, *spelling normalization*, *case folding*, *tokenizin*, dan *filtering*.

### 4.2.1 Analisis Deskriptif

Analisis deskriptif dilakukan untuk mengetahui gambaran umum mengenai data tanggapan tentang IndiHome yang telah didapatkan menggunakan teknik *scraping*. Aspek yang dilihat dalam analisis ini adalah jumlah ulasan yang masuk berdasarkan waktu dan perbandingan jumlah alasan dari pengguna yang dikategorikan menjadi dua kategori, yaitu ulasan positif dan ulasan negatif.



Gambar 4. 6 Jumlah Ulasan Pengguna *myIndiHome* di *Google Play*

Berdasarkan Gambar 4.6, pengambilan data ulasan selama periode 1 November sampai 15 Desember 2020 diperoleh hasil yang cukup fluktuatif. Ulasan terbanyak didapatkan pada minggu ketiga November yaitu pada tanggal 15-22 November 2020 dengan jumlah 499 Ulasan. Hal ini diperkirakan karena periode ini merupakan periode awal setelah tenggat pembayaran tagihan layanan IndiHome. Setelah proses pembayaran, *customer* berharap layanan yang didapatkan bisa lebih maksimal. Namun pada kenyataannya layanan yang didapatkan *customer* pada periode tersebut masih sama bahkan ada yang lebih buruk dari periode sebelumnya. Oleh karena itu banyak *customer* yang menuliskan keluhannya pada kolom komentar aplikasi *myIndiHome*.

#### 4.2.2 *Pre-processing Data*

Data dalam penelitian ini merupakan data ulasan yang diambil dari kolom komentar aplikasi *myIndiHome* pada situs *Google Play* yang memiliki berbagai macam gaya penulisan sehingga data yang diperoleh merupakan data yang tidak terstruktur. Oleh karena itu, data tersebut perlu diubah menjadi data yang lebih terstruktur sebelum dilakukan klasifikasi. Tahapan untuk mengubah data yang tidak terstruktur menjadi data yang lebih terstruktur disebut tahap *pre-processing*. Tahap *pre-processing* dilakukan dengan bantuan aplikasi R Studio versi 1.4.1103. Adapun langkah-langkah *pre-processing* yang akan dilakukan adalah:

1. *Translate* ulasan bahasa asing

2. *Spelling normalization*
3. *Case folding*
4. *Tokenizing*
5. *Filtering*

#### 4.2.2.1 *Translate Ulasan Bahasa Asing*

Tahap *translate* merupakan tahap mengubah ulasan yang berbahasa Inggris maupun campuran (Inggris dan Indonesia) akan diubah menjadi bahasa Indonesia. Hal ini dikarenakan terdapat beberapa ulasan yang menggunakan bahasa asing yang dianggap penting. Pada data ulasan *myIndiHome* terdapat sekitar 120 ulasan yang berbahasa campuran, dan tidak terdapat yang berbahasa Inggris secara penuh. Proses *translate* dapat dilihat pada Tabel 4.1

Tabel 4. 1 Proses *Translate*

<i>Input</i>	<i>Output</i>
<i>Smart solution</i> dalam mengakses segala kebutuhan hanya dalam satu sistem. Aplikasi oke banget, terlengkap dan terupdate memenuhi kebutuhan pelanggan. <i>Good Service</i> , Pertahankan!	Solusi cerdas dalam mengakses segala kebutuhan hanya dalam satu sistem. Aplikasi oke banget, terlengkap dan terupdate memenuhi kebutuhan pelanggan. Pelayanan yang baik, Pertahankan!

Adapun contoh *pseudo code* untuk proses *translate* seperti yang tertera pada tabel 4.2

Tabel 4. 2 *Pseudo Code Translate*

<i>Pseudo Code</i>	<b>Fungsi</b>
<code>docs &lt;- tm_map(docs, gsub, pattern="smart solution", replacement = "solusi cerdas")</code>	Mengubah kata " <i>smart solution</i> " menjadi "solusi cerdas".

#### 4.2.2.2 *Spelling Normalization*

Proses *spelling normalization* dilakukan untuk memperbaiki kata-kata yang disingkat maupun salah ejaan dengan bentuk berbeda namun memiliki maksud yang sama, sebagai contoh pada kata "tolong" memiliki beberapa bentuk

penulisan, seperti “tlg”, “tlong”, dan “tlng”. Berikut merupakan contoh proses *spelling normalization*:

Tabel 4. 3 Proses *Spelling*

<i>Input</i>	<i>Output</i>
Tagian tiba tiba naik drastis gila <b>gk</b> sesuai perjanjian pembayaran perbulan ini penipuan <b>tlong</b> tanggap komentar saya admin Indihome!!	Tagian tiba tiba naik drastis gila tidak sesuai perjanjian pembayaran perbulan ini penipuan tolong di tanggap komentar saya admin Indihome!!

Proses ini dijalankan menggunakan *Pseudo Code* seperti pada tabel 4.4

Tabel 4. 4 *Pseudo Code Spelling Normalization*

<i>Pseudo Code</i>	<b>Fungsi</b>
<code>docs &lt;- tm_map(docs, gsub, pattern="tlong", replacement = "tolong". "tolong")</code>	Mengubah kata “tlong” menjadi “tolong”.

#### 4.2.2.3 *Case Folding*

Pada proses *Case Folding*, terdapat beberapa tahapan yang dilakukan, seperti mengubah teks menjadi *lowercase*, menghapus angka, dan menghapus tanda baca.

##### 1. Mengubah Text Menjadi Huruf Kecil (*Lowecase*)

Proses ini dilakukan dengan mengubah huruf besar menjadi huruf kecil dengan tujuan agar huruf besar dan huruf kecil tidak terdeteksi memiliki arti yang berbeda. Perbedaan arti yang dimaksud adalah seperti kata “layanan” dan “Layanan”. Kedua kata tersebut memiliki arti yang sama, namun dideteksi oleh sistem komputer sebagai kata yang berbeda karena memiliki susunan huruf kecil dan besar yang berbeda. Adapun sebagai contoh kalimat ulasan pada Tabel 4.5 berikut:

Tabel 4. 5 Proses Mengubah Teks Menjadi Huruf Kecil (*Lowecase*)

<i>Input</i>	<i>Output</i>
Apaan dikit ngedown lambat banget Pake <b>L</b> ayanan 10 mbs tpi kecepatannya aj gk nyampe 1 <b>MB</b>	apaan dikit ngedown lambat banget pake layanan 10 mbs tpi kecepatannya aj gk nyampe 1 mb

*Pseudo code* yang digunakan untuk menjalankan proses tersebut seperti pada tabel 4.6.

Tabel 4. 6 *Pseudo Code* Mengubah Besar ke Huruf Kecil

<i>Pseudo Code</i>	Fungsi
<code>docs &lt;- tm_map(docs, content_transformer(tolower))</code>	Mengubah huruf besar menjadi huruf kecil pada dokumen.

## 2. Menghapus Angka dan Tanda Baca

Angka dan tanda baca dalam kalimat dianggap sebagai pemisah kata (delimiter) dan tidak memiliki pengaruh pada *text mining*, sehingga komponen tersebut perlu dihapus dalam setiap ulasan. Adapun sebagai contoh kalimat ulasan pada Tabel 4.7 berikut:

Tabel 4. 7 Proses Menghapus Angka dan Tanda Baca

<i>Input</i>	<i>Output</i>
Dlm bulan ini sdh 2 kali internet tdk berfungsi ! Tgl.6 dan tgl.13 Desember. Sampai hari ini belum juga berfungsi. Mhn percepatan perbaikannya.	Dlm bulan ini sdh kali internet tdk berfungsi Tgl dan tgl Desember. Sampai hari ini belum juga berfungsi. Mhn percepatan perbaikannya.

Untuk menjalankan proses ini digunakan *pseudo code* seperti pada Tabel 4.8.

Tabel 4. 8 *Pseudo Code* Menghapus Angka dan Tanda Baca

<i>Pseudo Code</i>	Fungsi
<code>docs &lt;- tm_map(docs, toSpace, "[[:punct:]]")</code>	Menghapus tanda baca.
<code>docs &lt;- tm_map(docs, toSpace, "[[:digit:]]")</code>	Menghapus angka.

### 4.2.2.4 *Tokenizing*

*Tokenizing* merupakan proses pemisahan teks ulasan menjadi kata per kata yang tidak saling berpengaruh (*independent*). Potongan kata tersebut dinamakan token, yaitu sebuah entitas yang mempunyai nilai dalam penyusunan matriks dokumen. *Tokenizing* berguna untuk mempermudah perhitungan frekuensi kemunculan kata dalam dokumen. Berikut merupakan contoh proses *tokenizing*:

Tabel 4. 9 Proses *Tokenizing*

<i>Input</i>	<i>Output</i>
sangat membantu dan puas dengan pelayanan my indihome	“sangat” “membantu” “dan” “puas” “dengan” “pelayanan” “my” “indihome”

Untuk menjalankan proses diatas digunakan *pseudo code* seperti pada tabel

Tabel 4. 10 *Pseudo Code Tokenizing*

<i>Pseudo Code</i>	<b>Fungsi</b>
list.kata = str_split(kalimat, '\\s+')	Mengubah kalimat menjadi potongan kata ( <i>tokenizing</i> ).

#### 4.2.2.5 *Filtering*

*Filtering* merupakan proses penyaringan atau pemilihan kata dalam dokumen. Tahap ini bertujuan untuk mengurangi kata dalam *corpus* menggunakan *stopwords*. *Stopwords* terdiri dari kata-kata yang apabila dihilangkan tidak mengurangi informasi yang terdapat pada kalimat. Kata-kata yang termasuk dalam *stopwords* biasanya terdiri dari kata ganti orang (aku, kamu, kalian), kata penghubung (atau, dan, kemudian), kata tanya (apa, dimana, bagaimana), dan kata lain yang tidak memiliki makna penting dalam penentuan bahasan dokumen. Proses ini dilakukan menggunakan *R Studio* dengan *library* “*tm*” yang digunakan untuk proses *text mining* dan “*wordcloud*” yang digunakan untuk analisis teks. Adapun contoh dan *pseudo code* proses *filtering* seperti yang tertera pada Tabel 4.11 dan Tabel 4.12 berikut:

Tabel 4. 11 Proses *Filtering*

<i>Input</i>	<i>Output</i>
“sangat” “membantu” “dan” “puas” “dengan” “pelayanan” “my” “indihome”	“membantu” “puas” “pelayanan” “my” “my” “indihome” “indihome”

Tabel 4. 12 *Pseudo Code Filtering*

<i>Pseudo Code</i>	<b>Fungsi</b>
stopwordindo = readLines("stopword.csv")	Memasukkan file daftar <i>stopword</i> yang telah dibuat.

<i>Pseudo Code</i>	<b>Fungsi</b>
docs <- tm_map(docs, removeWords, stopwordsindo)	Menghapus <i>stopword</i> yang terdapat pada dokumen

Setelah dilakukan proses *pre-processing*, jumlah data awal yang berjumlah 2.539 berubah menjadi 2.534 dikarenakan terdapat ulasan yang hanya berisi karakter simbol maupun emotikon.

### 4.2.3 Representasi Model

Representasi model dilakukan dengan membuat model dari data yang masih berbentuk kata-kata agar dapat diolah dan dihitung. Representasi dilakukan dengan melakukan perhitungan skor sentimen. Skor tersebut dihitung berdasarkan jumlah kata yang terdeteksi pada kamus *lexicon*. Setiap ulasan akan diberikan label sentimen berdasarkan skor sentimen yang telah didapatkan.

#### 4.2.3.1 Perhitungan Skor Sentimen

Terdapat dua metode yang umumnya digunakan oleh peneliti dalam menentukan suatu teks mengandung kalimat positif atau negatif. metode pertama yaitu dengan membaca isi teks secara keseluruhan, kemudian memberikan label secara manual. Metode ini memiliki presisi yang cukup tinggi namun sangat sulit diterapkan untuk data yang berjumlah sangat besar karena memakan banyak waktu (Azhar, 2017). Metode kedua yaitu dengan menerapkan algoritma yang dapat mengekstrak kalimat opini secara otomatis. Salah satu algoritma yang umum digunakan adalah *lexicon based*. *Lexicon based* dapat mengekstrak kalimat opini dengan presisi yang sangat tinggi (Azhar, 2017) dan cepat karena dilakukan secara otomatis. Oleh karena itu dalam penelitian ini, proses perhitungan dilakukan menggunakan metode *Lexicon Based*.

Metode *lexicon based* bekerja dengan cara membuat kamus kata opini (*lexicon*) terlebih dahulu. Kata-kata yang terdapat pada kamus tersebut akan digunakan untuk mengidentifikasi kata positif dan negatif pada suatu kalimat. Rumus umum yang digunakan untuk mendapatkan skor sentimen adalah sebagai berikut:

$$\text{Skor} = \text{Jumlah kata positif} - \text{Jumlah kata negatif} \quad (4.1)$$

Sebagai contoh perhitungan skor sentimen tertera pada tabel 4.13

Tabel 4. 13 Perhitungan Skor Setimen

Ulasan	Kata Positif	Kata Negatif
layan jaring lelet kecewa	cepat	lelet
gilir tagih cepat		kecewa
Jumlah	1	2

Berdasarkan ulasan pada tabel 4.7 diketahui bahwa terdapat 1 kata positif dan 2 kata negatif berdasarkan kamus *lexicon*. Kata positif yang terdeteksi yaitu “cepat”, sedangkan kata negatinya yaitu “lelet” dan “kecewa”. Oleh karena itu perhitungan skor sentimennya sebagai berikut:

$$\text{Skor} = 1 - 2 = -1$$

Dalam penelitian ini, perhitungan skor semtimen dilakukan dengan bantuan aplikasi RStudio dengan bantuan *library* “tm” yang umum digunakan dalam proses *text mining*. Adapun *pseudo code* yang digunakan dalam proses ini adalah sebagai berikut:

Tabel 4. 14 *Pseudo Code* Perhitungan Skor Sentimen

<i>Pseudo Code</i>	Fungsi
<code>library(tm)</code>	1. Menjalankan <i>packages</i> “tm” yang telah terinstal pada R.
<code>kalimat &lt;- read.csv("hasil_preprocessing.csv", header = TRUE)</code>	2. Membuka file yang akan dilakukan proses <i>skoring</i> .
<code>positif &lt;- scan("positive.txt", what = "character", comment.char = ";")</code>	3. Melakukan <i>scanning</i> daftar kata positif yang tersimpan dalam format <i>.txt file</i> .
<code>negatif &lt;- scan("negative.txt", what = "character", comment.char = ";")</code>	4. Melakukan <i>scanning</i> daftar kata negatif yang tersimpan dalam format <i>.txt file</i> .
<code>kata.positif &lt;- c(positif)</code> <code>kata.negatif &lt;- c(negatif)</code>	5. Mengidentifikasi variabel <code>kata.positif</code> dan <code>kata.negatif</code>

<i>Pseudo Code</i>	<b>Fungsi</b>
<pre> score.sentiment = function(kalimat, kata.positif, kata.negatif, .progress='none') { require(plyr) require(stringr) scores = lapply(kalimat, function(kalimat, kata.positif, kata.negatif) { kalimat = gsub('[:,punct:]]', '', kalimat) kalimat = gsub('[:,cntrl:]]', '', kalimat) kalimat = gsub('\\d+', '', kalimat) kalimat = tolower(kalimat) list.kata = str_split(kalimat, '\\s+') kata2 = unlist(list.kata)  positif.matches = match(kata2, kata.positif) negatif.matches = match(kata2, kata.negatif)  positif.matches = !is.na(positif.matches) negatif.matches = !is.na(negatif.matches) score = sum(positif.matches) - (sum(negatif.matches)) return(score) }, kata.positif, kata.negatif, .progress= .progress) </pre>	<p>6. Melakukan proses skoring menggunakan <i>function</i> dengan tahapan:</p> <ol style="list-style-type: none"> <li>Menjalankan <i>packages plyr</i> dan <i>stringr</i>.</li> <li>Menggabungkan setiap daftar inisial menjadi sebuah struktur data (<i>array</i>).</li> <li>Menghapus <i>noise</i> dan melakukan <i>case folding</i>.</li> <li>Mengubah kalimat menjadi potongan kata (<i>tokenizing</i>) dan menyederhanakan daftar kata.</li> <li>Mengidentifikasi kata positif dan kata negatif pada setiap potongan kata.</li> <li>Mengindikasi kata positif dan negatif ke dalam bentuk logika.</li> <li>Menghitung jumlah skor sentimen dan menyimpan skor ke dalam bentuk tabel.</li> </ol>

<i>Pseudo Code</i>	<b>Fungsi</b>
<pre>scores.df = data.frame(score=scores, Kalimat = kalimat) return(scores.df) } hasil = score.sentiment(kalimat\$Ulasan, kata.positif, kata.negatif) head(hasil) View(hasil)</pre>	7. Memanggil fungsi hasil skoring yang telah dibuat.

#### 4.2.3.2 Pelabelan Kelas Sentimen

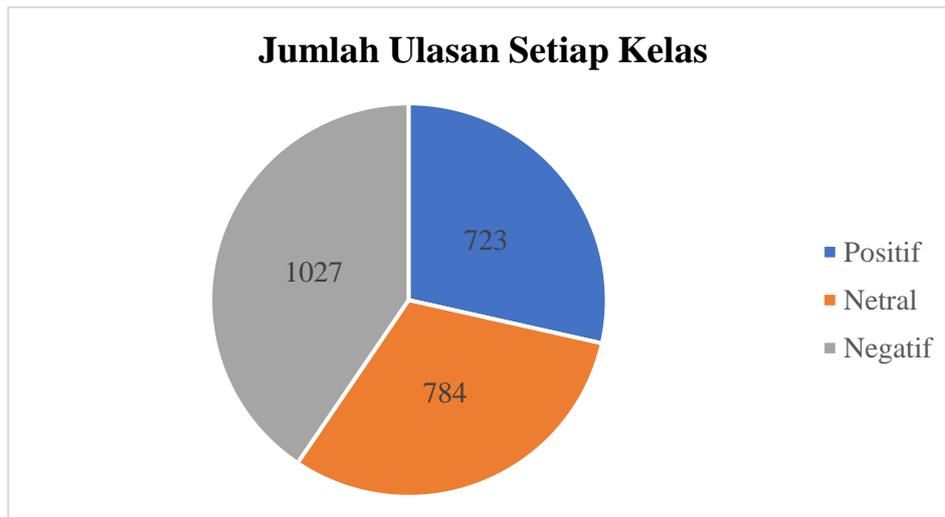
Setelah mendapatkan skor sentimen, pengolahan data dilanjutkan dengan melakukan pelabelan kelas setimen. Pelabelan dilakukan dengan membagi data ulasan menjadi tiga kelas sentimen, yaitu sentimen positif, netral dan negatif dengan ketentuan sebagai berikut (Buntoro, 2017):

- Sentimen negatif : skor < 0
- Sentimen netral : skor = 0
- Sentimen positif : skor > 0

Setelah melewati proses pembobotan dan pelabelan data ulasan menggunakan *software* R menggunakan *pseudo code* seperti pada tabel 4.15 dan didapatkan hasil perbandingan jumlah data dari pelabelan kelas setimen seperti yang tertera pada Gambar 4.7

Tabel 4. 15 *Pseudo Code* Pelabelan Kelas Sentimen

<i>Pseudo Code</i>	<b>Fungsi</b>
<pre>hasil\$klasifikasi &lt;- ifelse(hasil\$score&lt;0, "Negatif", "Positif") hasil\$klasifikasi data &lt;- hasil[c(3,1,2)] View(data) write.csv(data, "pelabelan.csv")</pre>	<p>1. Melakukan konversi nilai skor ke dalam kelas positif dan negaif</p> <p>2. Menyimpan file hasil pelabelan ke dalam format <i>.csv</i></p>



Gambar 4. 7 Diagram Jumlah Ulasan Setiap Kelas Sentimen

Berdasarkan Gambar 4.7 klasifikasi data dalam penelitian ini terbagi menjadi tiga kelas, yaitu positif sebanyak 723 ulasan, netral 789 ulasan, dan negatif 1.027 ulasan. Namun dalam penelitian ini hanya terdapat dua kelas yang digunakan, yaitu kelas positif dan negatif. Hal ini disebabkan karena kelas sentimen netral dinilai kurang memberi masukan dan manfaat bagi pihak perusahaan. Pada ulasan kelas sentimen positif mengandung pernyataan positif seperti ungkapan terimakasih, pujian, dan lain sebagainya. Untuk ulasan kelas sentimen negatif mengandung pernyataan negatif seperti ungkapan ketidakpuasan, cacian, penghinaan, dan sebagainya dimasukkan ke dalam sentimen negatif. Kelas sentimen netral dapat terjadi karena dua kemungkinan, yaitu:

1. Tidak terdapat kata sentimen pada data yang terdeteksi pada kamus
2. Jumlah skor kata positif sama dengan jumlah skor negatif

Adapun contoh hasil pelabelan data ulasan seperti pada tabel 4.16

Tabel 4. 16 Pelabelan Kelas Sentimen Pada Ulasan

Kelas Sentimen	Skor	Ulasan
Negatif	-2	parah bayar mahal jaringan kecewa
Positif	1	aplikasinya mantap
Netral	0	cek detail pelanggan layar putih tampilan apapun

Pada penelitian ini dilakukan reduksi kelas dengan mengkategorikan kelas sentimen ke dalam kelas sentimen netral positif dan negatif yang dilakukan

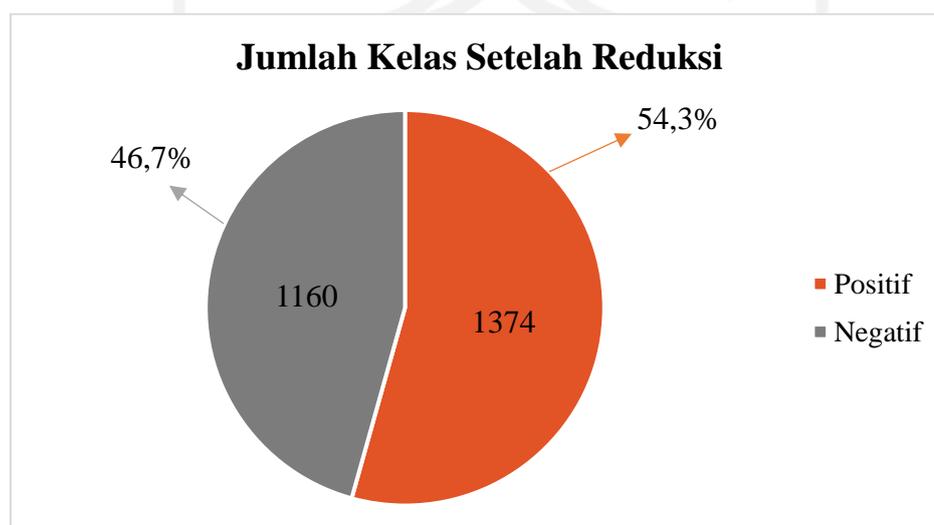
secara manual. Apabila kelas sentimen netral tidak teridentifikasi kata sentimen positif dan negatif maka akan dimasukkan ke dalam kelas sentimen positif dan apabila sentimen netral terdapat kata sentimen positif dan negatif yang seimbang maka akan dimasukkan ke dalam kelas sentimen negatif (Gumilang, 2018). Sebelumnya hal ini dilakukan dengan pertimbangan bahwa informasi negatif dapat diekstrak dengan lebih mudah untuk diterjemahkan sebagai keluhan atau ketidakpuasan pengguna, hal tersebut dapat menjadi masukan untuk pihak Telkom dapat melakukan perbaikan kearah yang lebih baik. Berikut merupakan contoh perhitungan skor sentimennya:

Tabel 4. 17 Perhitungan Skor Sentimen

Ulasan	Kata Negatif	Kata Positif
kualitas jaringanya <b>bagus</b> pertengahan <b>lemot</b> parah	lemot	bagus
<b>Jumlah</b>	<b>1</b>	<b>1</b>

$$\begin{aligned}
 \text{Skor} &= \text{jumlah kata positif} - \text{jumlah kata negatif} \\
 &= 1 - 1 \\
 &= 0
 \end{aligned}$$

Skor yang didapatkan dari simulasi perhitungan adalah 0 dan ditemukan kata sentimen positif dan negatif, maka ulasan akan dimasukkan ke dalam kelas negatif. Adapun hasil reduksi pelabelan kelas netral tertera pada Gambar 4.8.



Gambar 4. 8 Pelabelan Kelas Sentimen Setelah Reduksi

Berdasarkan Gambar 4.8 diperoleh hasil pelabelan kelas sentimen dengan ulasan positif memiliki frekuensi lebih banyak dibandingkan ulasan negatif. Dari total 2.539 ulasan, terdapat jumlah ulasan positif sebanyak 1.374 ulasan (54,22%) dan ulasan negatif sebanyak 1.160 ulasan (45,78%).

#### 4.2.4 Analisis Klasifikasi

Setelah melakukan tahap pelabelan kelas, proses pengolahan data dilakukan dengan analisis klasifikasi. Hasil data pelabelan dibagi menjadi dua, yaitu data *training* dan data *testing*. Proses klasifikasi menggunakan dua algoritma, yaitu *Support Vector Machine* (SVM) dan *Naïve Bayes Classifier* (NBC).

##### 4.2.4.1 Pembuatan Data *Training* dan Data *Testing*

Data *training* digunakan untuk membentuk model klasifikasi. Model ini merupakan representasi *knowledge* yang akan digunakan untuk prediksi kelas data baru. Sedangkan data *testing* digunakan untuk mengukur performa dari model yang telah didapatkan. Berdasarkan *Pareto Principle*, rasio yang umum digunakan untuk data *training* dan data *testing* adalah 80:20. Namun tidak menutup kemungkinan dalam sebuah penelitian tidak hanya menggunakan perbandingan tersebut. Hal tersebut dikarenakan jumlah data training akan mempengaruhi akurasi. Semakin banyak data training maka model akan semakin banyak belajar sehingga ketelitian akan semakin baik (Arrofiqoh & Harintaka, 2018).

Pada penelitian ini digunakan tiga nilai perbandingan data training dan testing. Berikut merupakan perbandingan jumlah data *training* dan data *testing*:

- 1) Perbandingan data training sebesar 70% dan data testing 30%:

Tabel 4. 18 Perbandingan Data *Training* dan Data *Testing* 1

Jenis Data	Presentase	Jumlah
Data Training	70%	1.773,8 $\approx$ 1.774
Data Testing	30%	760,2 $\approx$ 760
<b>Jumlah</b>	<b>100%</b>	<b>2.534</b>

Berdasarkan Tabel 4.10, terlihat bahwa perbandingan data *training* dan data *testing* sebesar 70% : 30% dari total 2.534 ulasan, terdapat sebanyak 1.774 ulasan sebagai data *training* dan 760 ulasan sebagai data *testing*.

- 2) Perbandingan data training sebesar 80% dan data testing 20%:

Tabel 4. 19 Perbandingan Data *Training* dan Data *Testing* 2

Jenis Data	Presentase	Jumlah
Data Training	80%	2.027,2 $\approx$ 2.027
Data Testing	20%	506,8 $\approx$ 507
<b>Jumlah</b>	<b>100%</b>	<b>2.534</b>

Berdasarkan Tabel 4.11, terlihat bahwa perbandingan data *training* dan data *testing* sebesar 80% : 20% dari total 2.534 ulasan, terdapat sebanyak 2.027 ulasan sebagai data *training* dan 507 ulasan sebagai data *testing*.

- 3) Perbandingan data training sebesar 90% dan data testing 10%:

Tabel 4. 20 Perbandingan Data *Training* dan Data *Testing* 3

Jenis Data	Presentase	Jumlah
Data Training	90%	2.280,6 $\approx$ 2.281
Data Testing	10%	253,4 $\approx$ 253
<b>Jumlah</b>	<b>100%</b>	<b>2.534</b>

Berdasarkan Tabel 4.12, terlihat bahwa perbandingan data *training* dan data *testing* sebesar 90% : 10% dari total 2.534 ulasan, terdapat sebanyak 2.281 ulasan sebagai data *training* dan 253 ulasan sebagai data *testing*.

#### 4.2.4.2 Klasifikasi dengan *Support Vector Machine* dan *Naïve Bayes Classifier*

Pada metode SVM, terdapat beberapa kernel yang digunakan, seperti kernel *Linear*, *Polynomial*, *Radial Basis Function* (RBF), dan *Sigmoid*. Dari setiap kernel akan dicari akurasi terbaik yang nantinya akan digunakan pada perbandingan akurasi dengan metode NBC. Berikut merupakan hasil perbandingan dari keempat kernel yang telah dilakukan pengujian:

Tabel 4. 21 Perbandingan Hasil Akurasi pada Klasifikasi SVM

Kernel	Akurasi
<i>Linear</i>	88,95%
<i>Polynomial</i>	64,10%
<i>RBF</i>	87,97%
<i>Sigmoid</i>	85,75%

Pada Tabel 4.21 dapat diketahui bahwa kernel *Linear* memiliki akurasi tertinggi dibandingkan kernel yang lainnya. Oleh karena itu, untuk proses klasifikasi selanjutnya akan digunakan kernel *Linear* dalam proses klasifikasi metode SVM.

Untuk menjalankan fungsi klasifikasi dengan algoritma SVM pada R Studio digunakan *pseudo code* seperti pada tabel 4.22

Tabel 4. 22 *Pseudo Code* Proses Klasifikasi Dengan Algoritma SVM

<i>Pseudo Code</i>	Fungsi
<code>df &lt;- read.csv("pelabelan.csv", stringsAsFactors = FALSE)</code>	1. Memasukkan data ke dalam R Studio.
<code>f&lt;- df[sample(nrow(df)),] glimpse(df) set.seed(10)</code>	2. Mengatur direktori kerja dan urutan dataset secara acak.
<code>Positiftraining= readLines("trainpos.csv") negatiftraining= readLines("trainneg.csv") positiftesting = readLines("testpos1.csv") negatiftesting = readLines("testneg1.csv")</code>	3. Membuka file data latihan dan data uji dalam format <i>csv</i> .
<code>reviewtraining = c(positiftraining, negatiftraining)</code>	4. Mendefinisikan masing-masing data latihan dan data uji.
<code>reviewtesting = c(positiftesting, negatiftesting)</code>	
<code>review_all = c(reviewtraining, reviewtesting)</code>	5. Menggabungkan data latihan dan data uji.
<code>sentiment_training = c(rep("positiftraining", length(positiftraining) ) ,rep("negatiftraining", length(negatiftraining)))</code>	6. Mendefinisikan label kelas pada data latihan.

<i>Pseudo Code</i>	<b>Fungsi</b>
<pre>sentiment_test = c(rep("positiftraining", length(positiftesting) ),rep("negatiftraining", length(negatiftesting))) sentiment_all = as.factor(c(sentiment_training, sentiment_test))  mat = create_matrix(review_all, language = "indonesian", removeStopwords = FALSE, removeNumbers = TRUE, stemWords = FALSE, tm::weightTf) mat = as.matrix(mat)  container &lt;- create_container(mat, sentiment_all, trainSize = 1:1774, testSize = 1775:2534, virgin = FALSE)  model &lt;- train_model(container, 'SVM', kernel = 'linear')  result &lt;- classify_model(container, model)  table(as.character(sentiment_all[1 775:2534]),  as.character(result[, "SVM_LABEL"]) )</pre>	<p>7. Mendefinisikan label kelas pada dat uji.</p> <p>8. Menggabungkan kelas dan mengubah data menjadi tipe data faktor.</p> <p>9. Membuat objek kelas kelas <i>DocumentTermMatrix</i>.</p> <p>10. Mengubah data dalam bentuk matriks.</p> <p>11. Membuat wadah untuk proses <i>training</i> dan <i>testing</i> data.</p> <p>12. Membuat model dari data <i>training</i> menggunakan algoritma SVM.</p> <p>13. Mengklasifikasikan data baru dari model yang telah didapatkan.</p> <p>14. Membuat tabel <i>confusion matrix</i>.</p>

<i>Pseudo Code</i>	<b>Fungsi</b>
<code>recall_accuracy(sentiment_all[1775:2534], result[, "SVM_LABEL"])</code>	15. Menampilkan nilai akurasi.
<code>create_precisionRecallSummary(cont ainer, result)</code>	16. Menampilkan nilai <i>precision</i> dan <i>recall</i> .

Adapun *pseudo code* proses klasifikasi dengan algoritma NBC seperti pada tabel 4.23.

Tabel 4. 23 *Pseudo Code* Proses Klasifikasi Dengan Algoritma NBC

<i>Pseudo Code</i>	<b>Fungsi</b>
<code>df &lt;- read.csv("pelabelan.csv", stringsAsFactors = FALSE)</code>	1. Memasukkan data ke dalam R Studio.
<code>set.seed(10)</code>	2. Mengatur direktori kerja dan urutan dataset secara acak
<code>df &lt;- df[sample(nrow(df)),]</code>	
<code>df &lt;- df[sample(nrow(df)),]</code>	
<code>glimpse(df)</code>	
<code>df\$class &lt;- as.factor(df\$class)</code>	3. Menjadikan kolom "class" sebagai label kelas.
<code>corpus &lt;- Corpus(VectorSource(df\$Kalimat))</code>	4. Melakukan <i>tokenization</i> dan menjadikan kolom "Kalimat" sebagai <i>corpus</i> .
<code>corpus</code>	
<code>inspect(corpus[1:3])</code>	
<code>corpus.clean &lt;- corpus %&gt;%</code>	5. Melakukan proses <i>case folding</i> dan <i>filtering</i> .
<code>tm_map(content_transformer(tolower)</code>	
<code>) %&gt;%</code>	
<code>tm_map(removePunctuation) %&gt;%</code>	
<code>tm_map(removeNumbers) %&gt;%</code>	
<code>tm_map(removeWords,</code>	
<code>stopwords(kind="en")) %&gt;%</code>	
<code>tm_map(stripWhitespace)</code>	

<i>Pseudo Code</i>	<b>Fungsi</b>
<pre>dtm &lt;- DocumentTermMatrix(corpus) inspect(dtm[40:50, 10:15])</pre>	6. Mengubah data dalam bentuk matriks.
<pre>df.train &lt;- df[1:1774,] df.test &lt;- df[1775:2534,] dtm.train &lt;- dtm[1:1774,] dtm.test &lt;- dtm[1775:2534,] corpus.train &lt;- corpus[1:1774] corpus.test &lt;- corpus[1775:2534]</pre>	7. Pembagian data <i>training</i> dan <i>testing</i> .
<pre>dim(dtm.train) fivefreq &lt;- findFreqTerms(dtm.train, 10) length((fivefreq)) fivefreq dtm.train.nb &lt;- DocumentTermMatrix(corpus.train, control = list(dictionary=fivefreq)) dim(dtm.train.nb) dtm.test.nb &lt;- DocumentTermMatrix(corpus.test, control = list(dictionary=fivefreq)) dim(dtm.train.nb)</pre>	8. Melakukan proses pemilihan fitur untuk memudahkan proses klasifikasi.
<pre>convert_count &lt;- function(x){   y &lt;- ifelse(x&gt;0,1,0)   y &lt;- factor(y, levels = c(0,1), labels = c("No","Yes"))   y} trainNB &lt;- apply(dtm.train.nb, 2, convert_count)</pre>	9. Mengubah data menjadi tipe data <i>boolean Naïve Bayes</i> .
	10. Perisapan proses klasifikasi.

<i>Pseudo Code</i>	<b>Fungsi</b>
<pre>testNB &lt;- apply(dtm.test.nb, 2, convert_count) naive=system.time(classifier &lt;- naiveBayes(trainNB, df.train\$class, laplace = 0)) print(naive) system.time (pred &lt;- predict(classifier, newdata = testNB))  conf.mat &lt;- confusionMatrix(table("Perdiction"= pred, "Actual"= df.test\$class)) conf.mat conf.mat\$byClass conf.mat\$overall conf.mat\$overall['Accuracy']</pre>	<p>11. Membuat model dari data <i>training</i> menggunakan algoritma NBC.</p> <p>12. Mengklasifikasikan data baru dari model yang telah didapatkan.</p> <p>13. Membuat dan menampilkan tabel <i>Confusion Matrix</i>.</p> <p>14. Menampilkan nilai <i>recall</i>, <i>precision</i>, dan akurasi.</p>

Proses klasifikasi dilakukan dengan membuat *confusion matrix* untuk mengetahui tingkat akurasi, *recall*, dan presisi. Matriks ini digunakan untuk mengevaluasi performansi model yang dibentuk oleh setiap algoritma klasifikasi. Dalam melakukan evaluasi model, dilakukan lima kali percobaan *dataset* untuk mendapatkan nilai akurasi terbaik (Putri, 2019).

Pada penelitian ini dilakukan lima kali percobaan untuk setiap data set. Perbedaan setiap percobaan terdapat pada berapa kali data set diacak dengan pembuatan bilangan *pseudorandom* banyak  $n$  kali menggunakan rumus *set.seed(n)* pada *software R*. Bilangan *pseudorandom* adalah bilangan yang dihasilkan melalui proses randomisasi yang semu. Semu yang dimaksud disini adalah bilangan *pseudorandom* bukanlah bilangan yang murni dihasilkan secara acak, melainkan melalui suatu proses yang memiliki suatu bahan atau bibit (*seed*). *Seed* yang dipilih akan dijadikan parameter dari fungsi *random*. Terdapat 5 *seed* yang digunakan dalam penelitian ini, yaitu 1 pada percobaan 1, 3 pada percobaan 2, 5 pada percobaan 3, 7 pada percobaan 4, dan 10 pada percobaan

5. Berikut merupakan rangkuman nilai akurasi setiap percobaan dalam penelitian ini:

Tabel 4. 24 Perbandingan Akurasi Model Klasifikasi SVM dan NBC

Percobaan	Akurasi Model	
	SVM	NBC
<b>Perbandingan 70:30</b>		
Percobaan 1	87,24%	84,61%
Percobaan 2	87,24%	84,34%
Percobaan 3	87,5%	83,95%
Percobaan 4	87,37%	83,68%
Percobaan 5	87,37%	84,47%
Rata-Rata	87,34%	84,21%
<b>Perbandingan 80:20</b>		
Percobaan 1	89,35%	86,59%
Percobaan 2	89,15%	84,81%
Percobaan 3	88,95%	84,22%
Percobaan 4	88,95%	84,81%
Percobaan 5	89,15%	82,84%
Rata-Rata	89,11%	84,65%
<b>Perbandingan 90:10</b>		
Percobaan 1	86,17%	90,11%
Percobaan 2	86,17%	83%
Percobaan 3	86,17%	84,98%
Percobaan 4	86,17%	85,38%
Percobaan 5	86,17%	82,6%
Rata-Rata	86,17%	85,2%
<b>Rata-rata Total</b>	<b>86,54%</b>	<b>84,69%</b>

Berdasarkan Tabel 4.24 diketahui bahwa nilai akurasi dengan metode SVM lebih besar dibandingkan metode NBC untuk semua percobaan kecuali pada percobaan pertama dengan perbandingan 90:10. Selain itu, nilai rata-rata total akurasi metode SVM lebih besar yaitu 86,54% dibandingkan metode NBC yang memiliki rata-rata total akurasi sebesar 84,69%. Oleh karena itu, dapat

disimpulkan bahwa algoritma SVM memiliki kinerja lebih baik dalam melakukan klasifikasi data ulasan *myIndiHome* dibandingkan metode NBC.

*Confusion matrix* digunakan untuk memudahkan dalam proses perhitungan akurasi dengan mengetahui jumlah data uji yang terklasifikasi dengan benar dan jumlah data uji yang salah pengklasifikasiannya.:

*Confusion matrix* menjelaskan seberapa besar presentase data yang diprediksi dengan benar oleh masing-masing percobaan. Jumlah data observasi berkategori negatif yang mampu diprediksi negatif (diprediksi dengan benar) disebut *true* negatif. Jumlah data observasi berkategori positif yang mampu diprediksi negatif (diprediksi dengan benar) disebut *true* positif. Banyaknya data observasi yang berkategori positif akan tetapi diprediksi negatif disebut dengan *false* positif. Banyaknya data observasi yang berkategori negatif akan tetapi diprediksi negatif disebut dengan *false* negatif.

Percobaan pertama pada kombinasi data 80:20 dengan metode SVM didapatkan prediksi bahwa pada kelas positif terdapat 287 ulasan yang diprediksi benar dan 27 ulasan yang diprediksi salah dari total ulasan positif sebesar 314 ulasan. Sedangkan pada kelas negatif, terdapat 166 ulasan yang diprediksi benar sebagai ulasan negatif dan 27 ulasan diprediksi salah sebagai ulasan positif dari total ulasan negatif sebanyak 193 ulasan. Dari nilai *confusion matrix* tersebut didapatkan nilai akurasi sebesar 89,35% yang memiliki arti bahwa dari 507 data ulasan yang diujikan, terdapat 453 ulasan yang diklasifikasikan benar oleh model dari algoritma *Support Vector Machine*.

Dengan menggunakan metode NBC pada prediksi kelas positif, didapatkan 254 ulasan benar diprediksi positif dan 36 ulasan salah dari total kelas positif sebanyak 290 data ulasan. Sedangkan pada kelas negatif, didapatkan sebanyak 185 data diprediksi benar sebagai kelas negatif dan 32 diprediksi salah dari total kelas negatif sebesar 217 data ulasan. Dari nilai *confusion matrix* tersebut diperoleh tingkat akurasi sebesar 86,59% yang artinya dari 507 data ulasan yang diujikan, terdapat 439 data ulasan yang diprediksi benar oleh model dari algoritma *Naïve Bayes Classifier*.

Adapun proses perhitungan nilai akurasi dilakukan dengan rumus sebagai berikut:

$$Akurasi = \frac{Jumlah\ prediksi\ benar}{Jumlah\ semua\ data\ uji} \times 100\% \quad (4.2)$$

Sehingga, untuk metode SVM nilai akurasi diperoleh dari perhitungan berikut:

$$Akurasi = \frac{166 + 287}{507} \times 100\%$$

$$Akurasi = \frac{453}{507} \times 100\%$$

$$Akurasi = 89,35\%$$

Sedangkan untuk nilai akurasi metode NBC didapatkan dari perhitungan berikut:

$$Akurasi = \frac{185 + 254}{507} \times 100\%$$

$$Akurasi = \frac{439}{507} \times 100\%$$

$$Akurasi = 86,59\%$$

Berikut merupakan *confusion matrix* kedua metode berdasarkan percobaan pertama dengan perbandingan data 80:20

Tabel 4. 25 *Confusion Matrix*

Prediksi	SVM		NBC	
	Negatif	Positif	Negatif	Positif
Negatif	166	27	185	36
Positif	27	287	32	254
<b>Akurasi</b>	<b>89,35%</b>		<b>86,59%</b>	

Berdasarkan perhitungan yang telah dilakukan, didapatkan bahwa performansi dari metode SVM lebih baik dari pada metode NBC. Hal tersebut ditunjukkan dengan nilai akurasi SVM lebih besar dari nilai akurasi NBC yaitu sebesar 89,35%.

#### 4.2.5 Visualisasi dan Asosiasi

Visualisasi dilakukan pada setiap klasifikasi kelas sentimen. Visualisasi dilakukan dengan tujuan untuk mengekstrasi informasi berupa topik yang paling sering dibicarakan/diulas oleh pengguna, sehingga dari keseluruhan teks yang ada dapat diambil informasi yang dianggap penting. Selain itu, perlu dilakukan mencari asosiasi kata yang paling sering muncul secara bersamaan. Proses ini digunakan untuk memperkuat informasi yang telah didapatkan dari proses

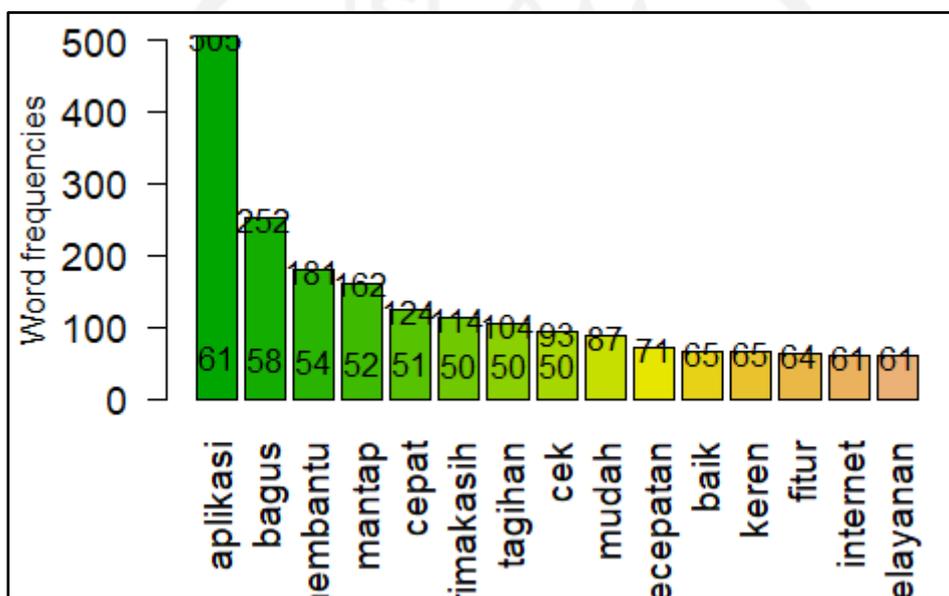
visualisasi. Terdapat beberapa *pseudocode* dalam melakukan proses visualisasi dan asosiasi seperti pada tabel 4.26.

Tabel 4. 26 *Pseudocode* Proses Visualisasi dan Asosiasi

<b><i>Barplot</i></b>	
<b><i>Pseudo Code</i></b>	<b>Fungsi</b>
<pre>k&lt;-barplot(d[1:15,]\$freq, las = 2, names.arg = d[1:15,]\$word,cex.axis=1.2,cex.names=1.2, main ="Most frequent words", ylab = "Word frequencies",col = terrain.colors(20))</pre>	1. Menampilk an <i>barplot</i> pada R Studio.
<b><i>Wordcloud</i></b>	
<b><i>Pseudo Code</i></b>	<b>Fungsi</b>
<pre>wordcloud(words = d\$word, freq = d\$freq, min.freq = 5, max.words=100, random.order=FALSE, rot.per=0.35, colors=brewer.pal(8, "Dark2"))</pre>	2. Menampilk an <i>wordcloud</i> pada R Studio.
<b><i>Asosiasi Kata</i></b>	
<b><i>Pseudo Code</i></b>	<b>Fungsi</b>
<pre>v&lt;-as.list(findAssocs(dtm, terms =c("aplikasi","bagus","membantu","mantap", "cepat", "terimakasih","tagihan","cek","mudah","kecep atan"), corlimit = c(0.13,0.07,0.13,0.02,0.17,0.13,0.17,0.17))) v</pre>	3. Menampilk an asosiasi kata yang 10 kata yang paling sering muncul.

#### 4.2.5.1 Ulasan Positif

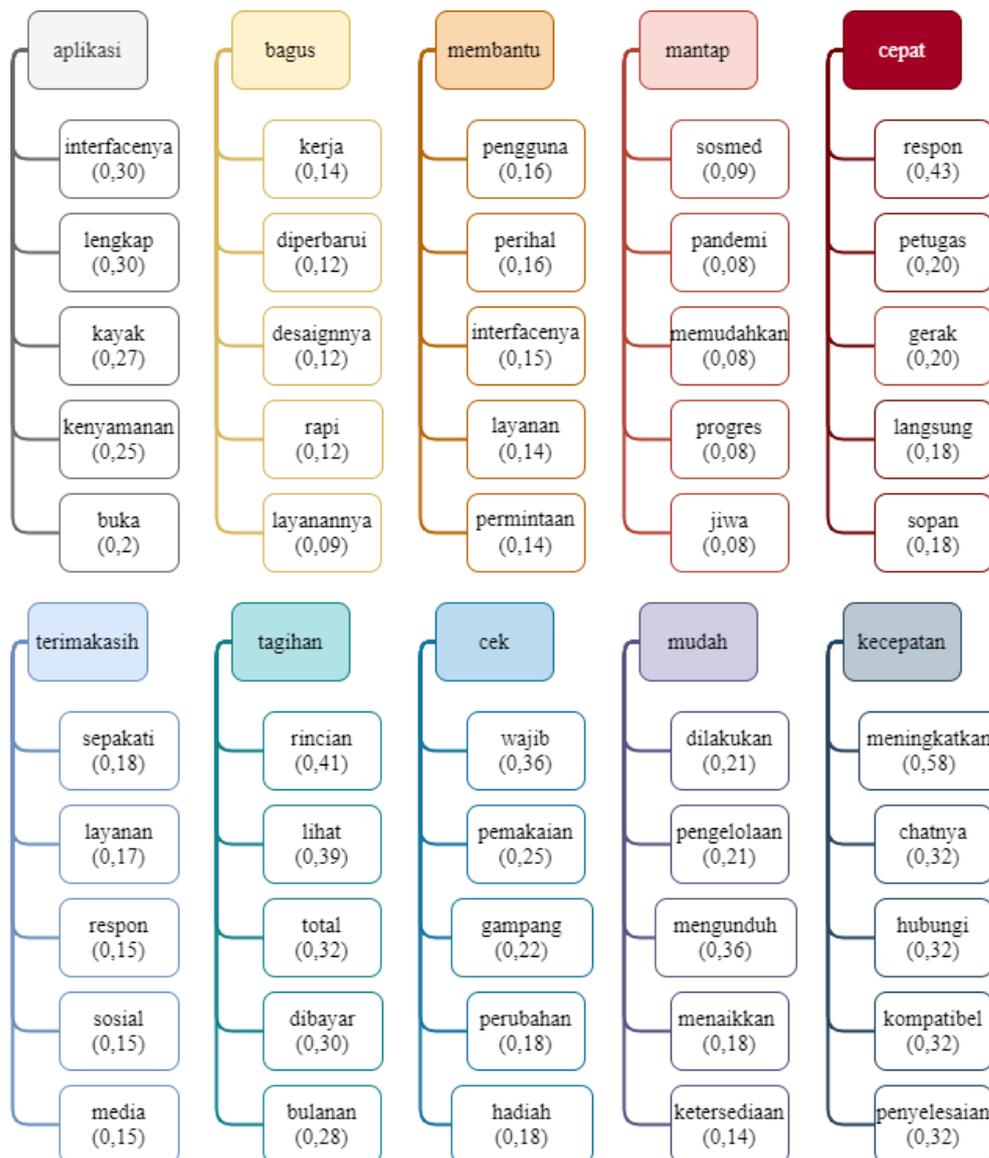
Data ulasan positif yang digunakan merupakan data hasil pelabelan yang dilakukan, baik menggunakan kamus *lexicon* maupun secara manual. Ekstraksi informasi pada ulasan positif dilakukan secara berulang-ulang hingga mendapatkan informasi tentang ulasan positif pengguna layanan *IndiHome* yang paling sering diulas/dibicarakan. Ulasan positif tersebut diidentifikasi berdasarkan frekuensi kata dalam ulasan. Berikut merupakan hasil visualisasi ekstraksi informasi yang didapatkan dari ulasan pengguna dengan klasifikasi ulasan positif.



Gambar 4. 9 Barplot Visualisasi Sentimen Positif

Pada hasil klasifikasi ulasan positif aplikasi *myIndiHome*, dari jumlah ulasan positif sebanyak 1.379 ulasan, diperoleh beberapa kata yang paling banyak muncul, seperti kata “aplikasi” dengan frekuensi 505 kali, “bagus” sebanyak 252 kali, “membantu” sebanyak 181 kali, dan seterusnya. Kata-kata yang muncul pada Gambar 4.9 merupakan kata yang memiliki sentimen positif dan merupakan topik pembicaraan yang paling sering diulas oleh pengguna *IndiHome*. Kata-kata tersebut selanjutnya digunakan sebagai dasar untuk menemukan asosiasi dengan kata lainnya, sehingga dapat diperoleh informasi yang lebih baik. Kumpulan kata-kata yang sering muncul tersebut dapat ditampilkan dalam bentuk *wordcloud* seperti berikut:





Gambar 4. 11 Asosiasi Kata Ulasan Sentimen Positif

Berdasarkan Gambar 4.11 diperoleh asosiasi kata pada klasifikasi kelas positif. Proses ekstraksi informasi dengan asosiasi dilakukan secara berulang-ulang dengan cara menyaring kata-kata yang memiliki hubungan dengan kata lain dan didasarkan pada relevansi kata dengan topik yang diulas.

Asosiasi kata yang berkaitan dengan kata “aplikasi”, dapat diperoleh informasi tentang aplikasi *myIndiHome* bahwa aplikasi memiliki *interface* yang bagus, lengkap, dan nyaman untuk digunakan ketika membukanya.

Kata-kata yang berasosiasi dengan kata “bagus” memberikan informasi bahwa setelah diperbarui, aplikasi memiliki desain yang bagus dan rapi. Selain itu, pengguna menilai karyawan telah melakukan kerja dengan bagus yang ditunjukkan dengan pelayanan yang diberikan.

Kata-kata yang berasosiasi dengan kata “membantu” memberikan informasi bahwa aplikasi telah membantu pengguna dalam melakukan permintaan pelayanan. Selain itu, *interface* aplikasi memudahkan pengguna dalam menggunakan aplikasi.

Kata-kata yang berasosiasi dengan kata “mantap” memberikan informasi bahwa aplikasi dapat memudahkan pelayanan pengguna selama pandemi, mulai dari pemasangan sampai progres pengaduan gangguan. Selain di aplikasi pengguna dapat melakukan pengaduan melalui sosial media seperti *twitter*.

Kata-kata yang berasosiasi dengan kata “cepat” memberikan informasi bahwa petugas merespon aduan pengguna dengan cepat. Setelah mendapatkan pengaduan, petugas langsung bergerak cepat untuk memberikan pelayanan terbaik. Selain itu, petugas juga bersikap sopan ketika memberikan pelayanan.

Kata-kata yang berasosiasi dengan kata “terimakasih” memberikan informasi bahwa pengguna mendapatkan layanan sesuai yang telah disepakati di awal. Layanan dapat disampaikan dan akan direspon oleh petugas melalui sosial media resmi *IndiHome*.

Kata-kata yang berasosiasi dengan kata “tagihan” memberikan informasi bahwa pengguna dapat melihat rincian dan total tagihan bulanan yang harus dibayar melalui aplikasi.

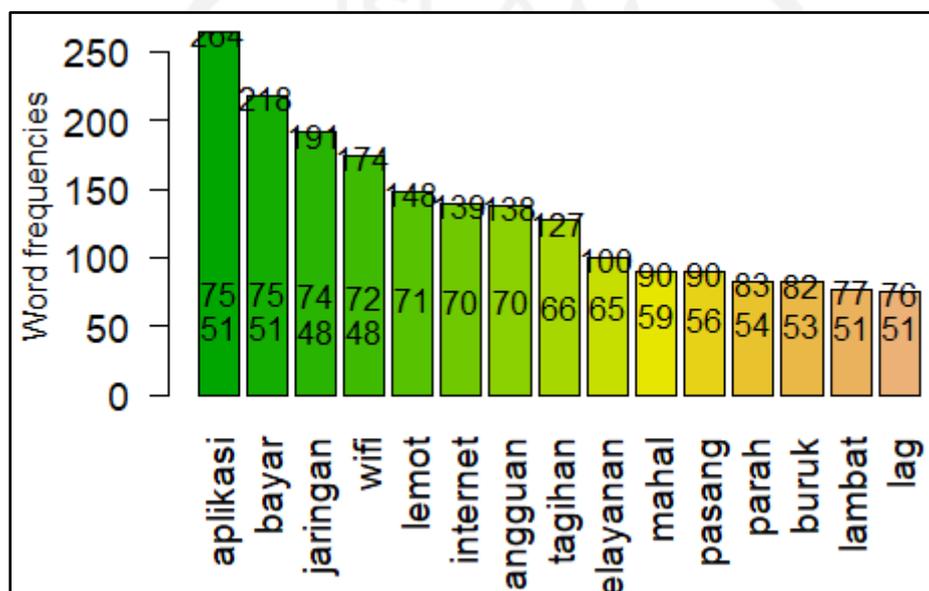
Kata-kata yang berasosiasi dengan kata “cek” memberikan informasi bahwa pengguna dapat mengecek pemakaian, perubahan layanan, dan hadiah secara mudah melalui aplikasi *myIndiHome*.

Kata-kata yang berasosiasi dengan kata “mudah” memberikan informasi bahwa aplikasi membantu pengguna untuk melakukan pengelolaan seperti pembayaran, menaikkan kecepatan, pengaduan apabila ada gangguan, dan mengecek ketersediaan layanan. Selain itu, layanan yang diberikan dapat mempermudah pengguna dalam mengunduh file-file yang dibutuhkan.

Kata-kata yang berasosiasi dengan kata “kecepatan” memberikan informasi bahwa kecepatan internet dapat meningkat dan kompatibel setelah pengguna mengajukan pengaduan.

#### 4.2.5.2 Ulasan Negatif

Ekstraksi informasi pada ulasan negatif dilakukan secara berulang-ulang hingga mendapatkan informasi tentang ulasan negatif pengguna *IndiHome* yang paling sering diulas/dibicarakan. Dari total ulasan sebanyak 2.534 ulasan, teridentifikasi sebanyak 1.160 ulasan negatif. Hasil ekstraksi informasi berupa ulasan negatif diidentifikasi berdasarkan frekuensi kata dalam ulasan, selain itu juga didasarkan pada relevansi kata dengan topik yang mengacu pada sentimen negatif. Berikut merupakan visualisasi hasil ekstraksi informasi yang didapatkan dari ulasan pengguna dengan klasifikasi negatif:



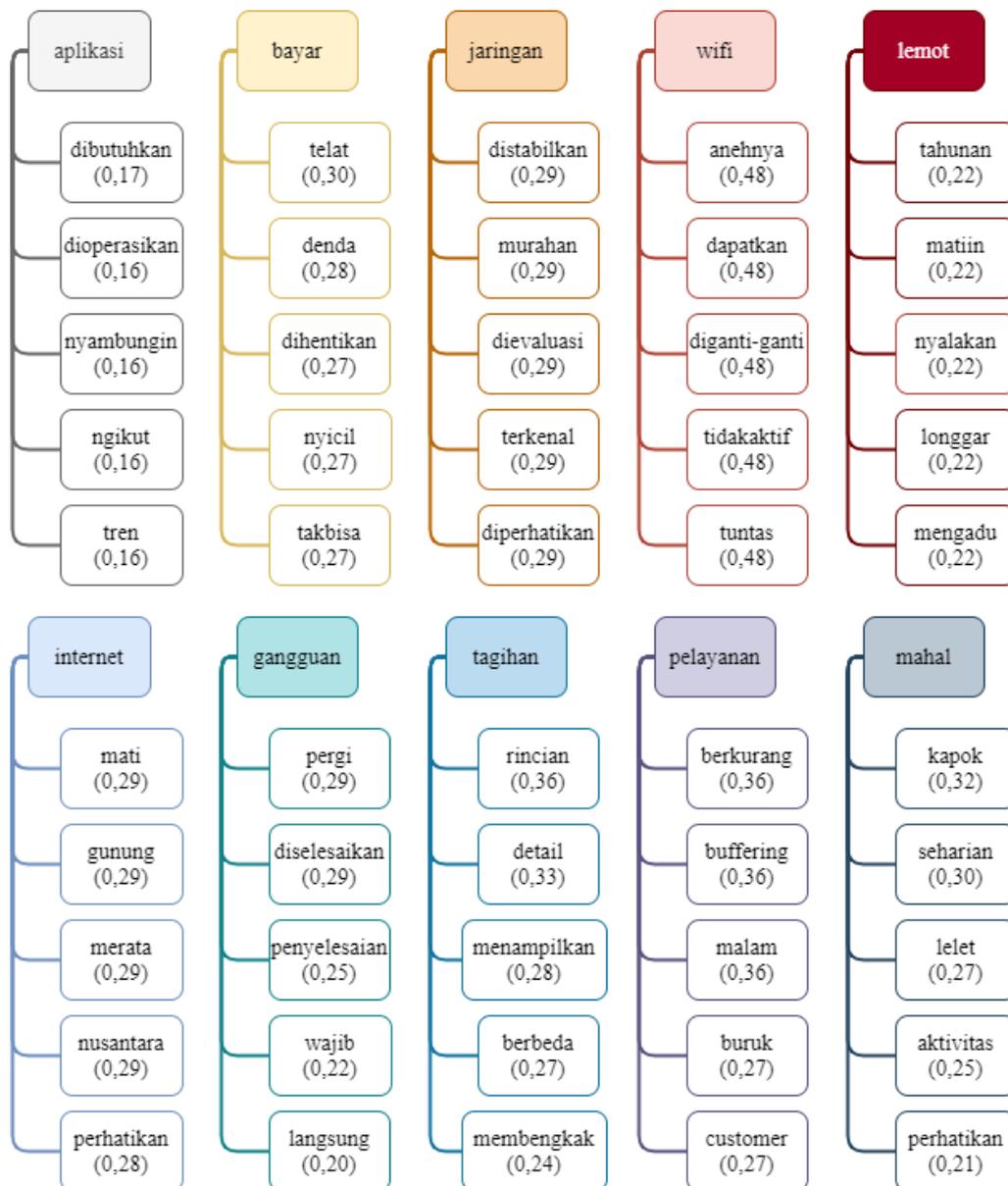
Gambar 4. 12 Barplot Visualisasi Sentimen Negatif

Pada hasil ulasan negatif diperoleh beberapa kata yang paling banyak muncul dengan topik yang dianggap relevan sebagai sentimen negatif, seperti kata “aplikasi” dengan frekuensi sebanyak 264 kali, “bayar” sebanyak 218 kali, “jaringan” sebanyak 191 kali, dan seterusnya. Kata-kata yang muncul seperti pada gambar 4.11 merupakan kata yang memiliki sentimen negatif dan merupakan topik pembicaraan yang paling banyak diulas oleh pengguna *IndiHome*. Kata-kata tersebut selanjutnya digunakan sebagai dasar untuk menemukan asosiasi kata yang lain, sehingga dapat diperoleh informasi berupa sentimen negatif yang lebih akurat. Kumpulan kata-kata yang sering muncul tersebut dapat ditampilkan dalam bentuk *wordcloud* seperti pada Gambar 4.12



Gambar 4. 13 *Wordcloud* Ulasan Sentimen Negatif

Visualisasi *wordcloud* pada Gambar 4.12 memberikan gambaran yang lebih jelas mengenai topik dan kata-kata negatif yang sering digunakan pengguna dalam memberikan ulasan. Beberapa topik yang sering dibahas pengguna diantaranya adalah mengenai aplikasi, bayar, jaringan, wifi, dan seterusnya. Kemudian, dilakukan pencarian asosiasi antar kata yang sering muncul secara bersamaan dan diperoleh hasil sebagai berikut:



Gambar 4. 14 Asosiasi Kata Ulasan Sentimen Negatif

Berdasarkan Gambar 4.13 diperoleh asosiasi kata yang menunjukkan asosiasi antar kata pada ulasan negatif. Kata-kata tersebut merupakan topik yang paling sering dibicarakan pengguna dalam ulasannya.

Kata-kata yang berasosiasi dengan kata “aplikasi” memberikan informasi bahwa aplikasi *myIndiHome* mejadi kebutuhan pengguna, namun ketika dioperasikan pengguna tidak bisa tersambung. Selain itu, *customer* menilai aplikasi yang dibuat hanya mengikuti *trend*. Hal tersebut dikarenakan aplikasi yang diberikan sering terjadi eror.

Kata-kata yang berasosiasi dengan kata “bayar” memberikan informasi bahwa apabila pengguna telat dalam melakukan pembayaran maka akan

dikenakan denda bahkan jaringan diputus seketika. Selain itu, apabila terdapat denda yang cukup membengkak, pengguna harus membayar secara kontan.

Kata-kata yang berasosiasi dengan kata “jaringan” memberikan informasi bahwa pengguna menilai jaringan yang didapatkan merupakan jaringan yang murahan dan tidak stabil. Pengguna memberikan saran untuk jaringan distabilkan, diperhatikan, dan dievaluasi karena *IndiHome* merupakan salah satu *brand* produk yang sudah terkenal.

Kata-kata yang berasosiasi dengan kata “wifi” memberikan informasi bahwa pengguna mendapatkan kecepatan wifi yang aneh yaitu berupa kecepatan wifi yang berganti-ganti atau tidak sesuai dengan yang dijanjikan. Tak jarang pula, wifi yang didapatkan tidak aktif.

Kata-kata yang berasosiasi dengan kata “lemot” memberikan informasi bahwa meskipun sudah menjadi langganan bertahun-tahun, kecepatan jaringan tetap lemot. Pengguna sudah mencoba untuk mematikan dan menyalakan kembali wifi, namun masih tetap lemot. Selain itu terdapat kabel yang longgar. Hal tersebut yang menjadikan pengguna sering mengajukan pengaduan.

Kata-kata yang berasosiasi dengan kata “internet” memberikan informasi bahwa internet sering mati pada lokasi pegunungan. Selain itu, layanan internet yang ditawarkan belum merata di seluruh nusantara. Hal ini menjadi perhatian penting bagi perusahaan.

Kata-kata yang berasosiasi dengan kata “gangguan” memberikan informasi bahwa pengguna harus pergi langsung ke plasa agar gangguan langsung diselesaikan. Hal tersebut dikarenakan aduan yang diajukan melalui aplikasi belum ada tanggapan.

Kata-kata yang berasosiasi dengan kata “tagihan” memberikan informasi bahwa pada aplikasi tidak menampilkan rincian dan detail tagihan. Hal tersebut berbeda dengan aplikasi versi sebelumnya. Selain itu pengguna sering mendapatkan tagihan yang membengkak tanpa ada rincian tagihan.

Kata-kata yang berasosiasi dengan kata “pelayanan” memberikan informasi bahwa pelayanan yang diberikan sudah berkurang yang ditunjukkan dengan jaringan yang buruk atau *buffering*. Selain itu pelayanan *customer* buruk ketika malam hari.

Kata-kata yang berasosiasi dengan kata “mahal” memberikan informasi bahwa pengguna merasa kapok menggunakan layanan *IndiHome*. Hal tersebut

dikarenakan pengguna sudah membayar mahal namun jaringan yang didapatkan selama sehari tetap lelet. Pengguna menyarankan untuk menjadikan hal ini sebagai perhatian penting dalam aktivitas perusahaan.



## BAB V PEMBAHASAN

### 5.1 Gambaran Umum Persepsi Pengguna *IndiHome*

Aplikasi *myIndiHome* di *Google Playstore* sudah diunduh sebanyak 5 juta lebih dengan dan 98 ribu ulasan (20/12/2020). Rating aplikasi ini hanya mencapai 3,6 dari 5 bintang. Nilai tersebut menunjukkan bahwa pelayanan yang diberikan PT Telkom pada layanan *IndiHome* belum sepenuhnya memenuhi harapan masyarakat.

Proses analisis sentimen yang telah dilakukan menunjukkan bahwa jumlah ulasan yang termasuk dalam kelas sentimen negatif sebanyak 1.160 ulasan dan kelas positif sebanyak 1.374 ulasan dari total 2.539 ulasan. Hasil tersebut menunjukkan bahwa *error* pelayanan dalam layanan *IndiHome* masih cukup tinggi, yaitu mencapai 46,7% yang ditunjukkan dengan jumlah ulasan negatif.

Berdasarkan penjelasan diatas, dapat didapatkan informasi bahwa hasil analisis sentimen relevan dan telah menginterpretasikan rating yang didapatkan aplikasi *myIndiHome*. Hasil analisis sentimen dan rating aplikasi *myIndiHome* menunjukkan bahwa *error* pelayanan pada layanan *IndiHome* masih sangat tinggi. Oleh karena itu perlu adanya evaluasi dan perbaikan kinerja pelayanan *IndiHome*.

### 5.2 Hasil Penerapan Metode *Support Vector Machine (SVM)* dan *Naïve Bayes Classifier (NBC)*

Proses klasifikasi diawali dengan membagi data uji dan data latih. Pada penelitian ini terdapat tiga perbandingan presentase data *training* dan data *testing* yang berbeda. Setiap bentuk perbandingan akan dilakukan percobaan sebanyak lima kali percobaan.

Pembagian data pertama dilakukan dengan data *training* 70% (1.774 ulasan) dan data *testing* 30% (760 ulasan). Dari lima kali percobaan, metode SVM memiliki akurasi terbaik pada percobaan ketiga yaitu sebesar 87% dengan nilai presisi kelas positif 88% dan negatif sebesar 87%. Sedangkan pada metode NBC, akurasi tertinggi didapatkan pada percobaan pertama yaitu sebesar 84,61% dengan presisi kelas positif sebesar 84,49% dan negatif sebesar 84,68%.

Berdasarkan rata-rata yang didapatkan, metode SVM memiliki performa lebih baik dengan rata-rata akurasi sebesar 87,34% dibandingkan metode NBC yang memiliki rata-rata akurasi sebesar 84,21%.

Pembagian data kedua dilakukan dengan data *training* 80% (2.027 ulasan) dan data *testing* 20% (507 ulasan). Berdasarkan hasil lima kali percobaan, metode SVM memiliki akurasi terbaik pada percobaan pertama yaitu mencapai 89,35% dengan nilai presisi kelas positif 91% dan negatif sebesar 86%. Pada metode NBC, akurasi tertinggi didapatkan pada percobaan pertama yaitu sebesar 86,59% dengan presisi kelas positif sebesar 88,81% dan negatif sebesar 83,71%. Berdasarkan rata-rata akurasi yang didapatkan, metode SVM memiliki performa lebih baik dengan nilai sebesar 89,11% dibandingkan metode NBC yang memiliki rata-rata akurasi sebesar 84,65%.

Pembagian data ketiga dilakukan dengan data *training* 90% (2.281 ulasan) dan data *testing* 10% (253 ulasan). Dari lima kali percobaan, Metode SVM memiliki nilai akurasi yang sama yaitu sebesar 86,17 dengan nilai presisi kelas positif 87% dan negatif sebesar 85%. Pada metode NBC, akurasi tertinggi didapatkan pada percobaan pertama yaitu sebesar 90,11% dengan presisi kelas positif sebesar 86,24% dan negatif sebesar 93,06%. Berdasarkan rata-rata akurasi yang didapatkan, metode SVM memiliki performa lebih baik dengan nilai sebesar 86,17% dibandingkan metode NBC yang memiliki rata-rata akurasi sebesar 85,2%.

Dari rata-rata akurasi lima kali percobaan didapatkan nilai akurasi metode SVM lebih tinggi dibandingkan metode NBC pada semua perbandingan data *training* dan *testing*. Selain itu, pada metode SVM rata-rata total akurasi lebih tinggi dengan nilai 86,54% dibandingkan dengan metode NBC yang memiliki nilai 84,69%. Oleh karena itu dapat disimpulkan bahwa metode SVM memiliki kemampuan yang lebih baik dalam mengklasifikasikan ulasan pengguna *IndiHome* dibandingkan metode NBC.

Nilai rata-rata total akurasi dari metode SVM dan NBC menunjukkan *error* lebih dari 13%, yang mana nilai ini harusnya dapat lebih diminimalkan. Nilai *error* ini dipengaruhi oleh jenis pelabelan kelas sentimen yang digunakan. Pendeteksian sentimen antar kata pada penelitian ini menggunakan kamus *lexicon* atau *bag of words*. Pada sistem ini dipelajari kata-kata yang terdapat dalam suatu sumber tanpa adanya pengkategorian kata berdasarkan struktur

penyusun kalimat, seperti kata benda, kata kerja, kata sifat, dan sebagainya. Hal tersebut yang menjadi pemicu terjadinya kesalahan pada sistem pelabelan oleh disebabkan oleh kerancuan padanan kata. Sebagai contoh kata “tidak memuaskan”, padanan kata ini memiliki kata “tidak” yang bersentimen negatif sedangkan kata “memuaskan” bersentimen positif, dimana keduanya bertolak belakang.

### 5.3 Hasil Klasifikasi dan Asosiasi Kata

Berdasarkan hasil klasifikasi ulasan *myIndiHome*, terdapat 1.379 ulasan positif. Dari ulasan tersebut, didapatkan beberapa kata yang paling sering muncul diantaranya adalah kata “aplikasi” dengan frekuensi sebanyak 505 kali, “bagus” sebanyak 252 kali, “membantu” sebanyak 181 kali, dan seterusnya seperti yang tertera pada Gambar 4.9. Kata-kata tersebut merupakan kata-kata yang memiliki sentimen positif dan menjadi topik pembicaraan yang paling banyak diulas oleh pengguna *IndiHome*.

Hasil klasifikasi ulasan menunjukkan bahwa terdapat 1.160 ulasan negatif. Dari ulasan tersebut, didapatkan beberapa kata yang paling sering muncul seperti kata “aplikasi” dengan frekuensi sebanyak 264 kali, “bayar” sebanyak 218 kali, “jaringan” sebanyak 191 kali, dan seterusnya seperti yang terlihat pada Gambar 4.12.

Kata-kata yang paling sering muncul dalam setiap kelas sentimen ini perlu dilakukan analisis asosiasi kata. Hal tersebut bertujuan agar peneliti mendapatkan informasi dan kesimpulan yang jelas atau tidak rancu.

Hasil asosiasi kata pada klasifikasi kelas sentimen positif seperti pada Gambar 4.11. menunjukkan proses ekstraksi informasi dengan asosiasi dilakukan secara berulang-ulang dengan cara menyaring kata-kata positif yang memiliki hubungan dengan kata lain dan didasarkan pada relevansi kata dengan topik yang diulas. Asosiasi kata yang diulas berkaitan dengan kata “aplikasi”, “bagus”, “membantu”, “mantap”, “cepat”, “terimakasih”, “tagihan”, “cek”, “mudah”, dan “kecepatan”. Pengguna *IndiHome* merasa aplikasi *myIndiHome* memiliki *interface* yang bagus, lengkap, dan nyaman untuk digunakan ketika membukanya. Aplikasi juga memiliki desain yang bagus dan rapi. Kelebihan-kelebihan aplikasi tersebut dinilai telah membantu pengguna dalam melakukan permintaan pelayanan seperti pengajuan pemasangan, menaikkan

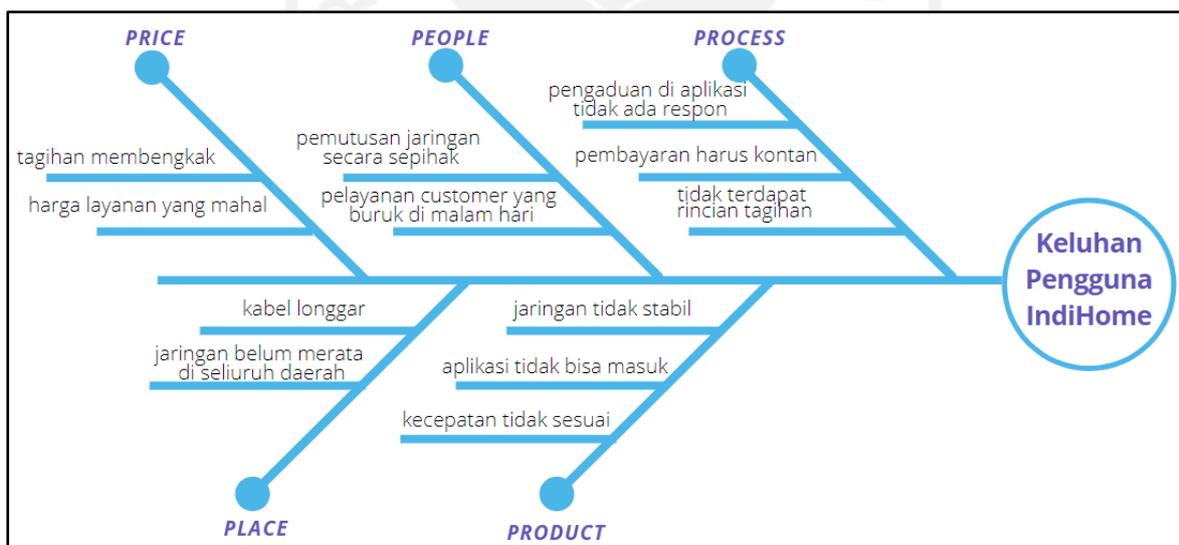
kecepatan, mengecek tagihan dan ketersediaan layanan sampai *progress* pengaduan gangguan jaringan. Dalam hal pelayanan, pengguna menilai karyawan telah melakukan kerja dengan bagus. Hal tersebut terlihat ketika petugas merespon aduan pengguna dengan cepat. Setelah mendapatkan pengaduan, petugas langsung bergerak cepat untuk memberikan pelayanan terbaik. Selain itu, petugas juga bersikap sopan ketika memberikan pelayanan. Selain itu, pengguna mendapatkan layanan sesuai yang telah disepakati di awal. Layanan dapat disampaikan dan akan direspon oleh petugas melalui sosial media resmi *IndiHome*. Dalam hal tagihan, pengguna dapat melihat rincian dan total tagihan bulanan yang harus dibayar melalui aplikasi. Dan juga pengguna dapat mengecek pemakaian, perubahan layanan, dan hadiah secara mudah melalui aplikasi *myIndiHome*. Adapun mengenai kecepatan internet, pengguna menilai kecepatan internet yang didapatkan meningkat dan kompatibel setelah pengguna mengajukan pengaduan. Hal tersebut memudahkan pengguna dalam mengunduh file-file yang dibutuhkan.

Hasil asosiasi kata pada klasifikasi kelas sentimen positif seperti pada Gambar 4.14. menunjukkan asosiasi antar kata pada ulasan negatif, yaitu “aplikasi”, “bayar”, “jaringan”, “wifi”, “lemot”, “internet”, “gangguan”, “tagihan”, “pelayanan”, “mahal”. Pengguna menilai aplikasi *myIndiHome* merupakan suatu kebutuhan, namun ketika dioperasikan pengguna tidak bisa tersambung. Selain itu, *customer* menilai aplikasi yang dibuat hanya mengikuti *trend*. Hal tersebut dikarenakan aplikasi yang diberikan sering terjadi eror. aplikasi ini tidak menampilkan rincian dan detail tagihan seperti aplikasi versi sebelumnya. Mengenai pembayaran, didapatkan informasi bahwa apabila pengguna telat dalam melakukan pembayaran maka akan dikenakan denda bahkan jaringan diputus seketika. Selain itu, apabila terdapat denda yang cukup membengkak, pengguna harus membayar secara kontan tanpa mendapatkan rincian tagihan. Mengenai jaringan internet, pengguna menilai jaringan yang didapatkan merupakan jaringan yang murahan, kecepatan tidak stabil dan tidak sesuai dengan yang dijanjikan. Meskipun sudah menjadi pelanggan bertahun-tahun, kecepatan jaringan yang didapatkan juga tetap lambat. Kendala tersebut sering terjadi karena beberapa hal, seperti permasalahan kabel, cuaca ekstrim dan wifi yang didapatkan tidak aktif. Pengguna memberikan saran untuk jaringan distabilkan, diperhatikan, dan dievaluasi karena *IndiHome* merupakan

salah satu *brand* produk yang sudah terkenal. Pengguna juga menilai layanan jaringan di pegunungan sering mati, dan layanan *IndiHome* belum merata di seluruh nusantara. Mengenai pelayanan yang diberikan, pengguna menilai pelayanan yang diberikan sudah berkurang yang ditunjukkan dengan jaringan yang buruk atau *buffering*. Selain itu pelayanan *customer* buruk ketika malam hari. Selain itu pengguna juga merasa kapok menggunakan layanan *IndiHome* karena pengguna sudah membayar mahal namun jaringan yang didapatkan selama sehari tetap lelet. Pengguna menyarankan untuk menjadikan hal ini sebagai perhatian penting dalam aktivitas perusahaan.

#### 5.4 Usulan Perbaikan Berdasarkan Ulasan Negatif Pengguna

Berdasarkan hasil ulasan negatif yang didapatkan, maka diperoleh informasi mengenai permasalahan yang terjadi terkait ulasan negatif pada *IndiHome* yang dianalisa dengan diagram *fishbone* berdasarkan faktor *price*, *people*, *process*, *place*, dan *product* sebagai berikut:



Gambar 5. 1 *Fishbone Diagram*

Berdasarkan diagram *fishbone* seperti Gambar 5.1 didapatkan informasi mengenai penyebab *myIndiHome* mendapatkan ulasan negatif dari pengguna. Terdapat lima faktor yang menjadi penyebabnya, yaitu faktor *price*, *people*, *process*, *place*, *product*. Selanjutnya perlu dilakukan penentuan langkah untuk pemecahan masalah yang ada. Berikut merupakan rencana pemecahan masalah pada layanan *IndiHome*:

Tabel 5. 1 Usulan Pemecahan Masalah

No.	Faktor	Permasalahan	Usulan Pemecahan Masalah
1.	<i>Price</i>	tagihan membengkak  Harga layanan yang mahal	Memberikan peringatan secara berkala ketika mendekati batas pembayaran.  Memberikan promo pada hari-hari besar, seperti natal dan tahun baru.
2.	<i>People</i>	Pemutusan jaringan secara sepihak  Pelayanan <i>customer</i> yang buruk di malam hari	Membuat SOP mengenai pemutusan jaringan, termasuk peringatan ketika akan dilakukan pemutusan jaringan.  Membuat sistem aduan 24 jam pada aplikasi.
3.	<i>Process</i>	Pengaduan di aplikasi tidak ada respon  Pembayaran harus kontan  Tidak terdapat rincian tagihan pada aplikasi	- Menambah karyawan <i>customer service</i> yang bertugas melayani <i>customer</i> via aplikasi.  - Pihak <i>developer</i> segera memperbaiki sistem aplikasi sehingga setiap aduan bisa diterima dan dibaca oleh petugas  Membuat sistem pembayaran yang dapat meringankan pengguna, seperti dapat dibayar secara bertahap.  Pihak <i>developer</i> segera memperbaiki sistem aplikasi sehingga rincian setiap tagihan dapat muncul di aplikasi dan diketahui pengguna.
4.	<i>Place</i>	Kabel longgar	- Melakuka <i>controlling</i> kualitas kabel secara berkala,

No.	Faktor	Permasalahan	Usulan Pemecahan Masalah
			<p>khususnya setelah cuaca ekstrim.</p> <ul style="list-style-type: none"> <li>- Segera menggan kabel yang telah tidak layak pakai.</li> </ul>
		Jaringan belum merata di seluruh daerah	<ul style="list-style-type: none"> <li>- Membuat program perluasan jaringan di setiap daerah.</li> <li>- Mengadakan promo khususnya kepada daerah yang masih jarang pengguna untuk mendorong pertumbuhan pengguna layanan.</li> </ul>
5.	Product	Jaringan tidak stabil	<ul style="list-style-type: none"> <li>- Memberikan pemberitahuan waktu apabila terdapat pemeliharaan, sehingga konsumen bisa memaklumi.</li> <li>- Melakukan <i>controlling</i> kualitas internet pada secara berkala, khususnya setelah cuaca buruk seperti hujan deras disertai angin.</li> </ul>
		Aplikasi tidak bisa masuk	<p>Pihak developer segera memperbaiki sistem aplikasi sehingga pengguna bisa masuk/login.</p>
		Kecepatan tidak sesuai.	<ul style="list-style-type: none"> <li>- Melakukan <i>controlling</i> secara berkala mengenai kecepatan yang didapatkan pengguna.</li> <li>- Melakukan tindakan tegas kepada pihak-pihak yang mencuri jaringan internet pengguna.</li> </ul>

## BAB VI PENUTUP

### 6.1 Kesimpulan

Berdasarkan analisis dan pembahasan yang telah dilakukan, dapat disimpulkan sebagai berikut:

1. Pada periode 1 Novemer – 15 Desember 2020 terdapat 2.539 ulasan aplikasi *myIndiHome* pada situs *Google Play*. Berdasarkan proses klasifikasi kelas sentimen didapatkan jumlah ulasan kelas positif sebanyak 1.379 ulasan (54,3%) dan ulasan negatif sebanyak 1.160 ulasan (46,7%).
2. Berdasarkan perbandingan data *training* dan *testing* sebesar 70%:30%, 80%:20%, 90%:10% dan dilakukan lima kali pengacakan *dataset* untuk setiap perbandingan didapatkan semua hasil akurasi menggunakan *Support Vector Machine* (SVM) lebih besar dari pada metode *Naïve Bayes Classifier* (NBC) kecuali pada percobaan pertama dengan perbandingan 90:10. Selain itu, nilai rata-rata total akurasi metode SVM lebih besar yaitu 87,54% dibandingkan metode NBC yang memiliki rata-rata total akurasi sebesar 84,69%. Oleh karena itu, dapat disimpulkan bahwa algoritma SVM memiliki kinerja lebih baik dalam melakukan klasifikasi data ulasan *myIndiHome* dibandingkan metode NBC.
3. Berdasarkan proses klasifikasi dan asosiasi kata yang dilakukan, didapatkan topik ulasan pengguna *IndiHome* yang sering dibicarakan baik pada kelas sentimen positif maupun negatif, seperti mengenai aplikasi, tagihan, pelayanan, kecepatan, jaringan, dan wifi. Pada kelas sentimen positif, pengguna sering membahas mengenai aplikasi, pelayanan, tagihan, dan kecepatan internet. Sedangkan pada kelas sentimen negatif, pengguna sering membahas mengenai aplikasi, pembayaran/tagihan, gangguan jaringan, wifi, kecepatan internet, pelayanan, dan mahal nya harga layanan.
4. Berdasarkan analisis pada *fishbone* diagram didapatkan 12 permasalahan hasil identifikasi pada ulasan negatif pengguna layanan *IndiHome* yang terbagi menjadi 5P faktor, yaitu *price*, *people*, *process*, *place*, dan *product*. Dua belas permasalahan tersebut yaitu tagihan membengkak, harga layanan yang mahal, pemutusan jaringan secara sepihak, pelayanan *customer* yang buruk ketika malam hari, pengaduan di aplikasi yang tidak direspon,

pembayaran harus secara kontan, tidak terdapat rincian tagihan pada aplikasi, kabel longgar, jaringan belum merata di seluruh daerah, jaringan tidak stabil, aplikasi tidak bisa masuk atau *login*, dan kecepatan yang tidak sesuai.

## 6.2 Saran

Berdasarkan hasil analisis, pembahasan, dan kesimpulan yang telah dilakukan, terdapat beberapa saran yang bisa diberikan yaitu:

1. Bagi pihak *IndiHome*

Hasil ekstraksi informasi dari ulasan-ulasan yang telah diberikan oleh pengguna, khususnya pada ulasan negatif dapat dijadikan sebagai bahan evaluasi dalam upaya peningkatan kepuasan pelanggan dan memberikan pelayanan semaksimal mungkin.

2. Bagi penelitian selanjutnya

Sebaiknya digunakan algoritma klasifikasi lain agar dapat diketahui perbandingan kinerja algoritma tersebut dibandingkan kedua algoritma yang telah digunakan, jumlah kombinasi dan pengacakan yang lebih banyak, serta periode pengumpulan data yang lebih panjang agar informasi yang diperoleh bisa lebih mendalam.

## DAFTAR PUSTAKA

- Alizah, M. D., Nugroho, A., Radiyah, U. & Gata, W., 2020. Sentimen Analisis Terkait "Lockdown" pada Sosial Media Twitter. *CSRID Journal* , 12(3), pp. 143-149.
- Al-Smadi, M. et al., 2018. Deep Recurrent Neural Network vs. Support Vector Machine for Aspect-based Sentiment Analysis of Arabic Hotels' Reviews. *Journal of Computational Science*, Volume 27, pp. 386-393.
- Alwasi'a, A., 2020. *Analisis Sentimen Pada Review Aplikasi Berita Online Menggunakan Metode Maximum Entropy (Studi Kasus: Review Detik.com pada Google Play 2019) [Skripsi]*. Yogyakarta: Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Islam Indonesia.
- Ananda, A. F. & Wandebori, H., 2016. The Impact of Drugstore Makeup Product Reviews by Beauty Vlogger on Youtube towards Purchase Intention by Undergraduate Students in Indonesia. *Journal International Conference on Ethics of Business, Economics, and Social Science*, pp. 264-273.
- Arifin, O. & Sasongko, T. B., 2018. Analisa Perbandingan Tingkat Performansi Metode Support Vector Machine dan Naive Bayes Classifier Untuk Klasifikasi Jalur Minat SMA. *Seminar Nasional Teknologi Informasi dan Multimedia*, 2(1), pp. 67-72.
- Arrofiqoh, E. N. & Harintaka, 2018. Implementasi Metode Convolutional Neural Network Untuk Klasifikasi Tanaman Pada Citra Resolusi Tinggi. *Geomatika*, 24(2), pp. 61-68.
- Asosiasi Penyelenggara Jasa Internet Indonesia, 2020. *Laporan Survey Internet APJII 2019-2020 (Q2)*. [Online] Available at:  
<https://apjii.or.id/survei2019x/download/BRTdltugLoJEG1wWiSrKZxeb0HDCQ5>  
 [Accessed 12 Desember 2020].
- Azhar, Y., 2017. Metode Lexicon-Learning Based Untuk Identifikasi Tweet Opini Berbahasa Indonesia. *Janapati*, 6(3), pp. 237-243.
- Baid, P., Gupta, A. & Chaplot, N., 2017. Sentiment Analysis of Movie Reviews Using Machine Learning Techniques. *International Journal of Computer Applications*, 179(7), pp. 45-49.

- Bordoloi, M. & Biswas, S., 2018. Sentiment Analysis of Product using Machine Learning Technique: A Comparison among NB, SVM and MaxEnt. *International Journal of Pure and Applied Mathematics*, 118(19), pp. 71-83.
- Budianto, A., Maryono, D. & Ariyuana, R., 2018. Perbandingan K-Nearest Neighbor (KNN) dan Support Vector Machine (SVM) Dalam Pengenalan Karakteristik Plat Kendaraan Kendaraan Bermotor. *Jurnal Ilmiah Pendidikan Teknik Kejuruan*, 11(1), pp. 1-9.
- Buntoro, G. A., 2017. Analisis Sentimen Calon Gubernur DKI Jakarta 2017 di Twitter. *Jurnal Integer*, Volume 1, p. 2.
- Clement, J., 2020. *Google Play: Number of Available Apps 2009-2020*. [Online] Available at: <https://www.statista.com/statistics/266210/number-of-available-applications-in-the-google-play-store/> [Accessed 15 January 2021].
- Dharmadhikari, S., Ingle, M. & Kulkarni, P., 2011. Empirical Studies on Machine Learning Based Text Classification Algorithms.. *Advanced Computing: An International Journal ( ACIJ )*, 6(2), pp. 161-169.
- Fadliyah, 2014. *Statistika: Terapannya di Informatika*. 1st ed. Yogyakarta: Graha Ilmu.
- Fanani, F., 2017. Klasifikasi Review Software Pada Google Play Menggunakan Pendekatan Analisis Sentimen. In: *Skripsi*. s.l.:Program Studi Teknologi Informasi Fakultas Teknik UGM Yogyakarta.
- Fanissa, S., Fauzi, M. A. & Adinugroho, S., 2017. Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode Naive Bayes dan Seleksi Fitur Query Expansion Ranking. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(8), pp. 2766-2770.
- Fritz, M., 2016. *Analisa Bad Hike Pada Kran Lavatory Tipe S11234R Menggunakan Metode Nominal Group Technique dan Metode Fishbone di PT Surya Toto Indonesia Tbk [Skripsi]*. Yogyakarta: Program Diploma Teknik Mesin Sekolah Vakasi Universitas Gajah Mada.
- Garcia, K. & Berton, L., 2020. Topic detection and sentiment analysis in Twitter content related to COVID-19 from Brazil and the USA. *Applied Soft Computing Journal*, pp. 1-25.

- Gumilang, Z., 2018. *Implementasi Naive-Bayes Classifier dan Asosiasi Untuk Analisis Sentimen Data Ulasan Aplikasi E-commerce Shopee pada Situs Google Play [Skripsi]*. Yogyakarta: Program Studi Statistika FMIPA Universitas Islam Indonesia.
- Handayani, A. S. et al., 2020. Klasifikasi Kualitas Udara Dengan Metode Support Vector Machine. *Jurnal Informatika & Rekayasa Elektronika*, 3(2), pp. 187-199.
- Han, J. & Kamber, M., 2006. *Data Mining: Concepts and Techniques*. 2nd ed. San Francisco: Morgan Kauffman.
- Han, J. & Kember, M., 2012. *Data Mining: Concepts and Techniques*. 3rd ed. Waltham, MA: Morgan Kaufmann.
- Harlian, M., 2006. *Machine Learning Text Kategorization*. Austin: University of Texas.
- Haryanto, D. J., Muflikhah, L. & Fauzi, M. A., 2018. Analisis Sentimen Review Barang Berbahasa Indonesia Dengan Metode Support Vector Machine Dan Query Expansion. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(9), pp. 2909-2916.
- Hearst, M., 2003. *What Is Text Mining?*. SIMS UC Berkeley Publication. [Online] Available at: <http://people.ischool.berkeley.edu/~hearst/text-mining.html> [Accessed 17 January 2021].
- Josi, A., Abdillah, L. A. & Suryayusra, 2014. Penerapan Teknik Web Scraping Pada Mesin Pencari Artikel Ilmiah. *Jurnal Sistem Informasi (SISFO)*, 5(1), pp. 159-164.
- Khammash, M., 2008. Electronic Word-of-Mouth: Antecedents of Reading Customer Reviews in On-line Opinion Platforms: A. *ADIS International Conference*.
- Manning, C. D., Raghavan, P. & Schutze, H., 2009. *An Introduction to Information Retrieval*. Cambridge: Cambridge University.
- Mohri, M., Rostamizadeh, A. & Talwalkar, A., 2012. *Foundations of Machine Learning*. Amazon: MIT Press.
- Mutawalli, L., Zaen, M. T. A. & Bagye, W., 2019. Klasifikasi Teks Sosial Media Twitter Menggunakan Support Vector Machine. *Jurnal Informatika & Rekayasa Elektronika*, 2(2), pp. 43-51.

- Muthia, D. A., 2017. Analisis Sentimen Pada Review Restoran Dengan Teks Bahasa Indonesia Menggunakan Algoritma Naive Bayes. *Jurnal Ilmu Pengetahuan dan Teknologi Komputer*, 2(2), pp. 2527-4864.
- Ngafifi, M., 2014. Kemajuan Teknologi dan Pola Hidup Manusia Dalam Prespektif Sosial Budaya. *Jurnal Pembangunan Pendidikan: Fondasi dan Aplikasi*, 2(1), pp. 33-47.
- Nurdin, M. D., 2016. *Dasar Sistem Telekomunikasi*. [Online] Available at: [https://repository.unimal.ac.id/1095/1/ST\\_01-Dasar-Sistem-Telekomunikasi.pdf](https://repository.unimal.ac.id/1095/1/ST_01-Dasar-Sistem-Telekomunikasi.pdf) [Accessed 15 January 2021].
- O'Keefe, T. & Koprinska, I., 2009. Feature Selection and Weighting Methods in Sentiment Analysis. *Proceedings of the 14th Australasian Document Computing Symposium*. Sydney, 4 December.
- Pang, B. & Lee, L., 2008. *Opinion Mining and Sentiment Analysis: Foundations and Trends*. [Online] Available at: <http://dx.doi.org/10.1561/1500000011> [Accessed 17 January 2021].
- Patel, R. & Passi, K., 2020. Sentiment Analysis on Twitter Data of World Cup Soccer Tournament Using Machine Learning. *IoT*, Volume 1, pp. 218-239.
- Permana, F. C., Rosmansyah, Y. & Abdullah, A. S., 2017. Naive Bayes as opinion classifier to evaluate students satisfaction based on student sentiment in Twitter Social Media. *Journal of Physics: Conference Series*, 7(6), pp. 1-9.
- Prasetyo, E., 2012. *Data Mining Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta: Andi Offset.
- PT Telekomunikasi Indonesia, 2019. *Laporan Tahunan 2019*. [Online] Available at: [https://telkom.co.id/sites/about-telkom/id\\_ID/page/ir-laporan-tahunan](https://telkom.co.id/sites/about-telkom/id_ID/page/ir-laporan-tahunan) [Accessed 20 Desember 2020].
- PT Telekomunikasi Indonesia, 2020. *Laporan Keuangan Telkom Kuartal III/2020, Telkom Mencatat Pertumbuhan Laba Bersih Positif di Tengah Pandemi*. [Online] Available at: [https://www.telkom.co.id/sites/about-telkom/id\\_ID/news/laporan-keuangan-telkom-kuartal-iii2020,-telkom-](https://www.telkom.co.id/sites/about-telkom/id_ID/news/laporan-keuangan-telkom-kuartal-iii2020,-telkom-)

[mencatat-pertumbuhan-laba-bersih-positif-di-tengah-pandemi-1211#:~:text=Pelanggan%20IndiHome%20tumbuh%20752%20ribu,pelanggan%20di%20akhir%20tahun%20ini.](#)

[Accessed 20 Desember 2020].

- Putri, B. A. D., 2019. *Analisis Sentimen Data Ulasan Pengguna Grab Menggunakan Metode Support Vector Machine dan Maximum Entropy. [Skripsi]*. Yogyakarta: Fakultas Teknologi Industri Universitas Islam Indonesia.
- Putri, D., 2016. *Implementasi Inferensi Fuzzy Mamdani Untuk Keperluan Sistem Rekomendasi Berita Berbasis Konten [Skripsi]*. Yogyakarta: Program Studi Ilmu Komputer FMIPA Universitas Gajah Mada.
- Rahmawati, A., Marjuni, A. & Zeniarja, J., 2017. Analisis Sentimen Publik Pada Media Sosial Twitter Terhadap Pelaksanaan Pilkada Serentak Menggunakan Algoritma Support Vector Machine. *Creative Communication and Innovative Technology Journal*, 10(2), pp. 197-2016.
- Ratino, Hafidz, N., Anggraeni, S. & Gata, W., 2020. Sentimen Analisis Informasi Covid-19 Menggunakan Support Vector Machine dan Naïve Bayes. *Jurnal Penelitian Ilmu dan Teknologi Komputer*, 12(2), pp. 1-11.
- Rofiqoh, U., Perdana, R. S. & Fauzi, M. A., 2017. Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan Lexicon Based Features. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1(12), pp. 1725-1732.
- Santoso, B., 2007. *Data Mining Teknik dan Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- Saraswati, N., 2011. *Text Mining dengan Metode Naive Bayes Classifier dan Support Vector Machines untuk Sentiment Analysis [Skripsi]*. Yogyakarta: Program Studi Teknologi Informasi Fakultas Teknik Universitas Gajah Mada.
- Setiawan, W., 2014. Sistem Pakar Diagnosa Penyakit Mata Menggunakan Naive Bayes Classifier. *Seminar Nasional Sains dan Teknologi*.
- Song, J. et al., 2017. A Novel Classification Approach Based on Naïve Bayes for Twitter Sentiment Analysis. *KSII Transactional On Internet And Information System*, 11(6), pp. 2996-3011.

- Steven, C., 2020. *Perbandingan Metode Analisis Sentimen Terhadap Pariwisata Indonesia Pada Media Sosial Twitter (Studi Kasus: Kota Bali) [Skripsi]*. Tangerang: Universitas Multimedia Nusantara.
- Sutanto, A. M. & Aprianingsih, A., 2016. The Effect of Online Consumer Review toward Purchase Intention: A Study in Premium Cosmetic in Indonesia. *Journal International Conference on Ethics of Business, Economics, and Social Science*, pp. 218-230.
- Syarifuddin, M., 2020. Analisis Sentimen Opini Publik Mengenai Covid-19 Pada Twitter Menggunakan Metode Naive Bayes dan KNN. *Inti Nusa Mandiri*, 15(1), pp. 23-28.
- Top Brand Award Indonesia, 2020. *Top Brand Index*. [Online] Available at: [https://www.topbrand-award.com/en/top-brand-index-int/?tbi\\_find=IndiHome](https://www.topbrand-award.com/en/top-brand-index-int/?tbi_find=IndiHome) [Accessed 20 Desember 2020].
- Turland, M., 2010. *Architect's Guide to Web Scraping Defined*. Los Angeles: Php.
- Ubaidillah, M. J., Munadhif, L. & Rinanto, N., 2019. Klasifikasi Gelombang Otot Lengan pada Robot Manipulator Menggunakan Support Vector Machine. *REKAYASA*, 2(12), pp. 91-97.
- Ulwan, M., 2016. *Pattern Recognition Pada Unstructured Data Teks Menggunakan Support Vector Machine dan Association [Skripsi]*. Yogyakarta: Program Studi Statistika FMIPA Universitas Islam Indonesia.
- Wati, R., 2016. Penerapan Algoritma Genetika Untuk Seleksi Fitur Pada Analisis Sentimen Review Jasa Maskapai Penerbangan Menggunakan Naive Bayes. *Jurnal Evolusi*, Volume 4.

## LAMPIRAN

### Lampiran 1. *Preprocessing*

```

install.packages("tm")
install.packages("wordcloud")
install.packages("RColorBrewer")

library("tm")
library("wordcloud")
library("RColorBrewer")
library(stringr)

View(preprocessing)
docs <- readLines("preprocessing.csv")

docs <- Corpus(VectorSource(docs))
inspect(docs)

toSpace <- content_transformer(function (x, pattern) gsub
(pattern, " ",x))
docs <- tm_map(docs,toSpace, "/" )
docs<-tm_map(docs,toSpace, "!")
docs<-tm_map(docs,toSpace, "@")

docs <- tm_map(docs, content_transformer(tolower))

docs <- tm_map(docs, toSpace, "[[:punct:]]")
docs <- tm_map(docs, toSpace, "[[:digit:]]")

docs <- tm_map(docs, gsub, pattern="nmr", replacement =
"nomor")
docs <- tm_map(docs, gsub, pattern="error", replacement =
"eror")
docs <- tm_map(docs, gsub, pattern="down", replacement =
"turun")
docs <- tm_map(docs, gsub, pattern="bener", replacement =
"benar")

```

```
docs <- tm_map(docs, gsub, pattern="login", replacement =
"masuk")
docs <- tm_map(docs, gsub, pattern="tdk", replacement =
"tidak")
docs <- tm_map(docs, gsub, pattern="yg", replacement = "yang")
docs <- tm_map(docs, gsub, pattern="dbayar", replacement =
"dibayar")
docs <- tm_map(docs, gsub, pattern="lgsg", replacement =
"langsung")
docs <- tm_map(docs, gsub, pattern="easy", replacement =
"mudah")
docs <- tm_map(docs, gsub, pattern="trs", replacement =
"terus")
docs <- tm_map(docs, gsub, pattern="gmn", replacement =
"bagaimana")
docs <- tm_map(docs, gsub, pattern="cpt", replacement =
"cepat")
docs <- tm_map(docs, gsub, pattern="bgt", replacement =
"banget")
docs <- tm_map(docs, gsub, pattern="dtg", replacement =
"datang")
docs <- tm_map(docs, gsub, pattern="problem", replacement =
"masalah")
docs <- tm_map(docs, gsub, pattern="tlg", replacement =
"tolong")
docs <- tm_map(docs, gsub, pattern="blom", replacement =
"belum")
docs <- tm_map(docs, gsub, pattern="useless", replacement =
"tidak berguna")
docs <- tm_map(docs, gsub, pattern="kmrn", replacement =
"kemarin")
docs <- tm_map(docs, gsub, pattern="kpd", replacement =
"kepada")

inspect(docs[1:10])
```

```
stopwordindo = readLines("stopword.csv")
docs <- tm_map(docs, removeWords, stopwordindo)
docs <- tm_map(docs, removeWords,
c("myindihome", "mbps", "indihome", "dan", "telkom", "mendapatkan",
kali", "saya", "gue", "aku", "ini", "semua", "buat", "kok", "aja", "nya"
))

docs <- tm_map(docs, removeURL)
removeURL <- function(x) gsub("http[[:alnum:]]*", "", x)

docs <- tm_map(docs, stripWhitespace)

inspect(docs)

dtm <- TermDocumentMatrix(docs)
m <- as.matrix(dtm)
v <- sort(rowSums(m), decreasing = TRUE)
d <- data.frame(word=names(v), freq=v)
head(d, 100)

dataframe <- data.frame(text=unlist(sapply(docs,
`[`) ), stringsAsFactors=F)
write.csv(dataframe, "hasil_preprocessing.csv")
```

## Lampiran 2. Pembobotan dan Pelabelan

```

library(tm)
kalimat <- read.csv("hasil_preprocessing.csv", header = TRUE)

#Scoring data berdasarkan kamus data
positif <- scan("positive.txt", what = "character",
comment.char = ";")
negatif <- scan("negative.txt", what = "character",
comment.char = ";")
kata.positif <- c(positif)
kata.negatif <- c(negatif)
score.sentiment = function(kalimat, kata.positif, kata.negatif,
                           .progress='none')
{
  require(plyr)
  require(stringr)
  scores = laply(kalimat, function(kalimat, kata.positif,
kata.negatif)
  {
    kalimat = gsub('[:punct:]', '', kalimat)
    kalimat = gsub('[:cntrl:]', '', kalimat)
    kalimat = gsub('\\d+', '', kalimat)
    kalimat = tolower(kalimat)

    list.kata = str_split(kalimat, '\\s+')
    kata2 = unlist(list.kata)
    positif.matches = match(kata2, kata.positif)
    negatif.matches = match(kata2, kata.negatif)
    positif.matches = !is.na(positif.matches)
    negatif.matches = !is.na(negatif.matches)
    score = sum(positif.matches) - (sum(negatif.matches))
    return(score)
  }, kata.positif, kata.negatif, .progress= .progress)
  scores.df = data.frame(score=scores, Kalimat = kalimat)
  return(scores.df)
}

```

```
hasil = score.sentiment(kalimat$Ulasan, kata.positif,
kata.negatif)
head(hasil)
View(hasil)

#Convert score to sentiment
hasil$klasifikasi <- ifelse(hasil$score<0, "Negatif",
"Positif")
hasil$klasifikasi
View(hasil)

#Tukar Row
data <- hasil[c(3,1,2)]
View(data)
write.csv(data, "pelabelan.csv")

#Menyimpan data positif dan negatif
data.pos <- hasil[hasil$score>0,]
View(data.pos)
write.csv(data.pos, "positif.csv")
data.neg <- hasil[hasil$score<0,]
View(data.neg)
write.csv(data.neg, "negatif.csv")
```

**Lampiran 3.** Klasifikasi Dengan Metode NBC

```

#Install Packages
install.packages("devtools")
install.packages("MIAMaxent")
install.packages("RTextTools")
install.packages("tm")
install.packages("doMC", repos="http://R-Forge.R-
project.org")
library(dev)

#Load required libraries
library(tm)
library(SparseM)
library(RTextTools)
library(e1071)
library(stats)
library(dplyr)
library(NLP)
library(caret)
library(MIAMaxent)
library(readr)

#library for parallel processing
library(doMC)
registerDoMC(cores = detectCores())

#input data
df <- read.csv("pelabelan.csv", stringsAsFactors = FALSE)
View(df)
set.seed(10)
df <- df[sample(nrow(df)),]
df <- df[sample(nrow(df)),]
glimpse(df)

```

```

df$class <- as.factor(df$class)
corpus <- Corpus(VectorSource(df$Kalimat))
corpus
inspect(corpus[1:3])

corpus.clean <- corpus %>%
  tm_map(content_transformer(tolower)) %>%
  tm_map(removePunctuation) %>%
  tm_map(removeNumbers) %>%
  tm_map(removeWords, stopwords(kind="en")) %>%
  tm_map(stripWhitespace)

dtm <- DocumentTermMatrix(corpus)
inspect(dtm[40:50, 10:15])

#Pembagian Training Set 70% dan Testing Set 30%
df.train <- df[1:1774,]
df.test <- df[1775:2534,]
dtm.train <- dtm[1:1774,]
dtm.test <- dtm[1775:2534,]
corpus.train <- corpus[1:1774]
corpus.test <- corpus[1775:2534]

dim(dtm.train)
fivefreq <- findFreqTerms(dtm.train, 10)
length((fivefreq))
fivefreq
dtm.train.nb <- DocumentTermMatrix(corpus.train, control
= list(dictionary=fivefreq))
dim(dtm.train.nb)
dtm.test.nb <- DocumentTermMatrix(corpus.test, control =
list(dictionary=fivefreq))
dim(dtm.train.nb)

```

```
#Fungsi Pelabelan
convert_count <- function(x){
  y <- ifelse(x>0,1,0)
  y <- factor(y, levels = c(0,1), labels = c("No","Yes"))
  y
}
trainNB <- apply(dtm.train.nb, 2, convert_count)
testNB <- apply(dtm.test.nb, 2, convert_count)

#Naive Bayes
install.packages("naivebayes")
library(naivebayes)
naive=system.time(classifier <- naiveBayes(trainNB,
df.train$class, laplace = 0))
print(naive)
system.time (pred <- predict(classifier, newdata =
testNB))

#Confusion Matrix
conf.mat <- confusionMatrix(table("Perdiction"=pred,
"Actual"= df.test$class))
conf.mat$byClass
conf.mat$overall
conf.mat$overall['Accuracy']
conf.mat
```

#### Lampiran 4. Klasifikasi Dengan Metode SVM

```

#Import data
df <- read.csv("pelabelan.csv", stringsAsFactors = FALSE)
f<- df[sample(nrow(df)),]
glimpse(df)
set.seed(10)

positiftraining = readLines("trainpos1.csv")
negatiftraining = readLines("trainneg1.csv")
positiftesting = readLines("testpos1.csv")
negatiftesting = readLines("testneg1.csv")

reviewtraining = c(positiftraining, negatiftraining)
reviewtesting = c(positiftesting, negatiftesting)

review_all = c(reviewtraining, reviewtesting)

sentiment_training = c(rep("positiftraining",
length(positiftraining) ),rep("negatiftraining",
length(negatiftraining)))
sentiment_test = c(rep("positiftraining",
length(positiftesting) ),rep("negatiftraining",
length(negatiftesting)))

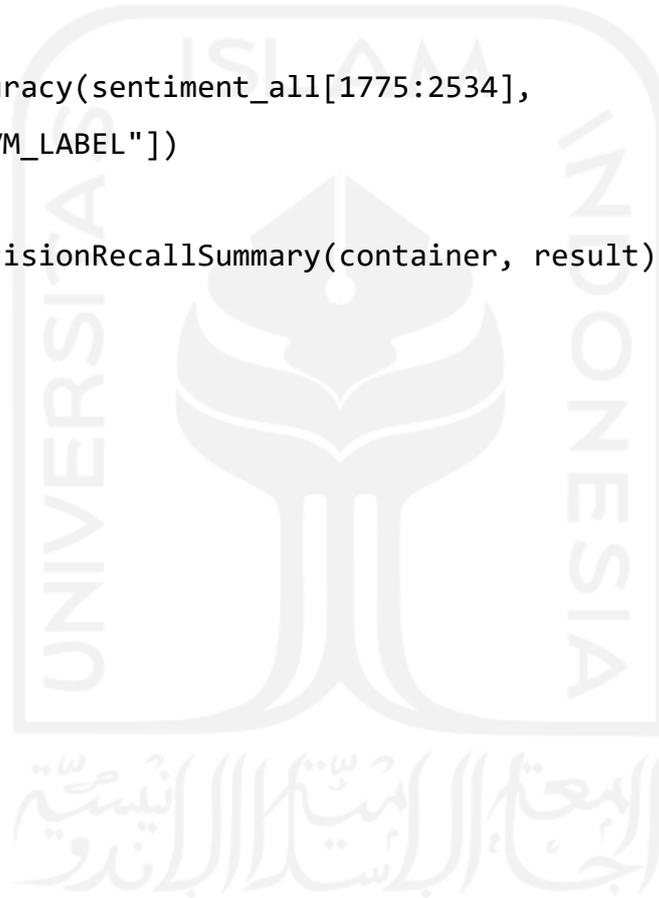
sentiment_all = as.factor(c(sentiment_training,
sentiment_test))

mat = create_matrix(review_all, language = "indonesian",
removeStopwords = FALSE, removeNumbers = TRUE, stemWords
= FALSE, tm::weightTf)

mat = as.matrix(mat)

```

```
container <- create_container(mat, sentiment_all,  
trainSize = 1:1774, testSize = 1775:2534, virgin = FALSE)  
  
model <- train_model(container, 'SVM', kernel = 'linear')  
result <- classify_model(container, model)  
  
table (as.character(sentiment_all[1775:2534]),  
       as.character(result[,"SVM_LABEL"]))  
  
recall_accuracy(sentiment_all[1775:2534],  
               result[,"SVM_LABEL"])  
  
create_precisionRecallSummary(container, result)
```



**Lampiran 3.** Visualisasi dan Asosiasi Kata

```
#Install package
install.packages("wordcloud2")

#Load Package
library(NLP)
library(tm)
library(SnowballC)
library(RColorBrewer)
library(wordcloud)
library(stringr)
library(wordcloud2)

#Input data sentimen positif
docs <- read.csv("positif.csv")
View(positif)

#Memanggil data sebagai Corpus
docs <- Corpus(VectorSource(docs$Kalimat))

#Menghilangkan stopwords
docs <- tm_map(docs,
removeWords,c("pakai","gak","banget","bantu","tingkat","b
anget","tolong","udah"))

docs <- tm_map(docs, stripWhitespace)
inspect(docs)

#Membangun Term Document Matrix
dtm<- TermDocumentMatrix(docs)
m <- as.matrix(dtm)
v <- sort(rowSums(m),decreasing=TRUE)
d <- data.frame(word = names(v),freq=v)
```

```

head(d, 100)

#WordCloud
set.seed(1234)
wordcloud(words = d$word, freq = d$freq, min.freq = 5,
           max.words=100, random.order=FALSE,
           rot.per=0.35,
           colors=brewer.pal(8, "Dark2"))

#Melihat frekuensi Kata
findFreqTerms(dtm, lowfreq = 4)

#Asosiasi Kata
v<-as.list(findAssocs(dtm, terms
=c("aplikasi","bagus","membantu","mantap", "cepat",
"terimakasih","tagihan","cek","mudah","kecepatan"),
           corlimit =
c(0.13,0.07,0.13,0.02,0.17,0.13,0.17,0.17)))
v

k<-barplot(d[1:20,]$freq, las = 2, names.arg =
           d[1:20,]$word,cex.axis=1.2,cex.names=1.2,
           main ="Most frequent words",
           ylab = "Word frequencies",col
           =topo.colors(20))

#Barplot
k<-barplot(d[1:15,]$freq, las = 2, names.arg =
           d[1:15,]$word,cex.axis=1.2,cex.names=1.2,
           main ="Most frequent words",
           ylab = "Word frequencies",col =
           terrain.colors(20))
termFrequency<- rowSums(as.matrix(dtm))

```

```
termFrequency<- subset(termFrequency, termFrequency>=48)
text(k,sort(termFrequency, decreasing = T)-
      1,labels=sort(termFrequency, decreasing = T),pch =
6, cex = 1)
```

